

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 03:18:03 ; Search time 10034 Seconds

(without alignments)
11717.693 Million cell updates/sec

Title: US-10-081-408-1

Perfect score: 4040
Sequence: 1 gtcctcccccacccttagtc.....ccccaaaaaaaaaaaaa 4040Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_dat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_ov:*

22: em_ox:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_mam:*

37: em_htg_vtc:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4039.6	100.0	4040	6	AX329732	AX329732 Sequence
2	4039.6	100.0	4040	6	AX330309	AX330309 Sequence
3	4039.6	100.0	4040	9	HSU9447	U99447 Human Place
4	2501	61.9	2501	9	AF067406	AF067406 Homo sapi
5	2277	56.4	2277	9	AK025727	AK025727 Homo sapi
6	1933.2	47.9	3624	4	AB019242	AB019242 Bos tauru
7	1853.2	45.9	189149	2	AC046171	AC046171 Homo sapi
8	1853.2	45.9	194567	2	AC055866	AC055866 Homo sapi
9	1849	45.8	3233	9	AB050500S3	AB050502 Homo sapi
10	1787.6	44.2	2605	4	BTY15774	Y15774 Bos tauru
11	1779.2	44.0	1794	9	AF054985	AF054985 Homo sapi
12	1761	43.6	2614	9	AB050500S1	AB050500 Homo sapi
13	1754	43.4	2664	4	BOYERA	L27218 Bos tauru
14	1754	43.4	2664	4	BOYERA	S69583 copper amin
15	1705.6	42.2	4210	10	AF115411	AF115411 Mus muscu
16	1702.4	42.1	4452	10	AF054831	AF054831 Mus muscu
17	1432.6	35.5	2611	4	AB022282	AB022282 Bos tauru
18	1359.2	33.6	2073	9	AF047485	AF047485 Homo sapi
19	1198.8	29.7	195444	2	AC025424	AC025424 Mus muscu
20	1198.8	29.7	221184	2	AL732315	AL732315 Mus muscu
21	1198.8	29.7	250353	10	AL590969	AL590969 Mouse DNA
22	1197.2	29.6	14357	10	AF078705	AF078705 Mus muscu
23	1146.6	28.4	2653	9	AF081363	AF081363 Homo sapi
24	1017.8	25.2	2572	9	D88213	D88213 Homo sapien
25	773.8	19.2	3741	9	AB012942S1	AB012942 Homo sapi
26	714.4	17.7	1051	6	AX401797	AX401797 Sequence
27	714.4	17.7	1051	6	RNU72632	U72632 Rattus norv
28	701.6	17.4	59015	2	AC067811	AC067811 Homo sapi
29	422.2	10.5	2441	6	AX330273	AX330273 Sequence
30	422.2	10.5	2441	6	AX335503	AX335503 Sequence
31	422.2	10.5	2441	9	HSU11862	U11862 Human clone
32	419	10.4	2455	9	BC014093	BC014093 Homo sapi
33	396.2	9.8	2668	10	RMABPL	X73911 R.norvegicu
34	393.2	9.7	2375	9	HUMABPA	M55602 Human amilo
35	393.2	9.7	2375	11	G18279	G18279 SWS1267 Er
36	388	9.6	2249	9	AK092514	AK092514 Homo sapi
37	381.8	9.5	2662	10	BC034215	BC034215 Mus muscu
38	380.2	9.4	2729	10	BC021880	BC021880 Mus muscu
39	355.2	8.8	2473	9	HSU11863	U11863 Human clone
40	337.2	8.3	1708	10	BC022627	BC022627 Mus muscu
41	320.4	7.9	360	6	AX017501	AX017501 Sequence
42	293	7.3	1001	9	AB050500S2	AB050501 Homo sapi
43	263.2	6.5	271	9	HUM170YACX	L77596 Homo sapien
44	259.2	6.4	991	9	AF047486	AF047486 Homo sapi
45	245.2	6.1	1721	9	AF218035	AF218035 Homo sapi

ALIGNMENTS

RESULT 1	AX329732	LOCUS	AX329732	4040 bp	DNA	linear	PAT 09-JAN-2002
DEFINITION	Sequence	241 from Patent WO0194629.					
ACCESSION	AX329732						
VERSION	AX329732.1	GI:18102710					
KEYWORDS							
SOURCE		human.					
ORGANISM		Homo sapiens					
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
AUTHORS		Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.					
TITLE		Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,					
		Horrigan,S., Soppet,D.R. and Weaver,Z.					
		Cancer gene determination and therapeutic screening using signature					

gene sets
Patent: WO 0194629-A 241 13-DEC-2001:
Avalon Pharmaceuticals (US)
Location/Qualifiers

FEATURES
source 1. .4040
/organism="Homo sapiens"
/db.xref="taxon:9606"

BASE COUNT 825 a 1183 c 1097 g 934 t 1 others
ORIGIN

Query Match 100.0%; Score 4039.6; DB 6; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTCTTCCACCCCTTAGTCCAGGATCTGACTACCGGGAACCTTAGCCAGAGTCCGGGA 60
Db 1 GTCTTCCACCCCTTAGTCCAGGATCTGACTACCGGGAACCTTAGCCAGAGTCCGGGA 60
QY 61 GCCCCCAACCCCGTCCAGAGGACCAAGAGCCCGCTGTGCTGGCGTGAATACATTG 120
Db 61 GCCCCCAACCCCGTCCAGAGGACCAAGAGCCCGCTGTGCTGGCGTGAATACATTG 120
QY 121 CTCTCTTTGGTGAATGAGCTGTCCCTTTCGTGGAAAAATGAACAGAACATCTCT 180
Db 121 CTCTCTTTGGTGAATGAGCTGTCCCTTTCGTGGAAAAATGAACAGAACATCTCT 180
QY 181 CGTGTCTCATTTCTGGCGGCTCATACCATCTTTGCCCTTGGTTGTCTCTGCTGGG 240
Db 181 CGTGTCTCATTTCTGGCGGCTCATACCATCTTTGCCCTTGGTTGTCTCTGCTGGG 240
QY 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTTCCCATTCGCCCTGTATCTCCAG 300
Db 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTTCCCATTCGCCCTGTATCTCCAG 300
QY 301 TGCCAGCCTTGGACACACCTTGCCAGAGCCAGCTTTTGCAGACTGAGCCGAGAGA 360
Db 301 TGCCAGCCTTGGACACACCTTGCCAGAGCCAGCTTTTGCAGACTGAGCCGAGAGA 360
QY 361 GCTGACGCTGGATGAGGCTTTCTGACCAAGCGGCTGGGGCCAGGGGCTGTGATGCA 420
Db 361 GCTGACGCTGGATGAGGCTTTCTGACCAAGCGGCTGGGGCCAGGGGCTGTGATGCA 420
QY 421 CCAGGCGCGGCGCTGGAACACTGTCTTCTCATGTGAGTGGAGTCCCTCCCAAGGC 480
Db 421 CCAGGCGCGGCGCTGGAACACTGTCTTCTCATGTGAGTGGAGTCCCTCCCAAGGC 480
QY 481 TGACGCTTGGCTCACTTGGACAGGGGAGGCCCCACCTGCGCGGAGGCACTGGCCT 540
Db 481 TGACGCTTGGCTCACTTGGACAGGGGAGGCCCCACCTGCGCGGAGGCACTGGCCT 540
QY 541 CGTCTTCTTTGGGAGGCAACCCCAAGCAAGTGAAGTGTGTGGGCGCACTGCC 600
Db 541 CGTCTTCTTTGGGAGGCAACCCCAAGCAAGTGAAGTGTGTGGGCGCACTGCC 600
QY 601 TCACCCCTCTCATGAGGAGTGAAGTGTGAGGCTCATGGAAGGCCCGCTCCCTATCA 660
Db 601 TCACCCCTCTCATGAGGAGTGAAGTGTGAGGCTCATGGAAGGCCCGCTCCCTATCA 660
QY 661 CCGAGCGCCCGCTGTCTCAAGAGTACCTGACATAGACAGATGATCTTCAACAGAGA 720
Db 661 CCGAGCGCCCGCTGTCTCAAGAGTACCTGACATAGACAGATGATCTTCAACAGAGA 720
QY 721 GCTGGCCCAAGGCTTGTGGGCTTCTCCACACCTGTGCTTCTCAAGAGCAGCGGGAGGAA 780
Db 721 GCTGGCCCAAGGCTTGTGGGCTTCTCCACACCTGTGCTTCTCAAGAGCAGCGGGAGGAA 780
QY 781 CCTGTGACAAATGACAGAGGCTCCCGTGTGTGCAATCAGAGGAGCCGAGCCACTGGTT 840
Db 781 CCTGTGACAAATGACAGAGGCTCCCGTGTGTGCAATCAGAGGAGCCGAGCCACTGGTT 840
QY 841 TGGCTTCACTACAACATCTGGGGCTGGGGTCTTCTCTGACCAAGTGGGCTTGGAGCT 900
Db 841 TGGCTTCACTACAACATCTGGGGCTGGGGTCTTCTCTGACCAAGTGGGCTTGGAGCT 900

```

```

QY 901 GCTAGTGAACCAAGGCCCTTGAACCTTGCCCGCTGAGCTATCCAGAAAGTGTCTATCA 960
Db 901 GCTAGTGAACCAAGGCCCTTGAACCTTGCCCGCTGAGCTATCCAGAAAGTGTCTATCA 960
QY 961 AGCGGCTACATACGACACGCTGGGCGCCAGCTGGAGGCCAGTTTGGGGCGGGCTGTAA 1020
Db 961 AGCGGCTACATACGACACGCTGGGCGCCAGCTGGAGGCCAGTTTGGGGCGGGCTGTAA 1020
QY 1021 TGTGTGCTGATCCAGACAAATGGCAGAGTGGGCTGTGCTCCCTGAAGTCCCTGTGCC 1080
Db 1021 TGTGTGCTGATCCAGACAAATGGCAGAGTGGGCTGTGCTCCCTGAAGTCCCTGTGCC 1080
QY 1081 CCCGGGTCCAGCTCCCGCTCTACATTTCTATCCCAAGGCCCGCTTCAGTGTCCAGGG 1140
Db 1081 CCCGGGTCCAGCTCCCGCTCTACATTTCTATCCCAAGGCCCGCTTCAGTGTCCAGGG 1140
QY 1141 AAGTGAAGGCGCTCTACAGTGTGAGCTTTCCTTGGCCCGGAGATTCAGTGGGCC 1200
Db 1141 AAGTGAAGTGGCTCTCTACATGTGAGCTTTCCTTGGCCCGGAGATTCAGTGGGCC 1200
QY 1201 AAGGATCTTTCAGCTTCCTTCCAGAGAGAAAGACTAGTTTATGAGATTAAGCTCCAGA 1260
Db 1201 AAGGATCTTTCAGCTTCCTTCCAGAGAGAAAGACTAGTTTATGAGATTAAGCTCCAGA 1260
QY 1261 GGCTTGGCCATCTATGGTGGAAATTCGCCAGCAGCAATGACAGCCCTATGTGATGG 1320
Db 1261 GGCTTGGCCATCTATGGTGGAAATTCGCCAGCAGCAATGACAGCCCTATGTGATGG 1320
QY 1321 AGGCTTTGGCATGGGCAAGTATACACAGCCCTGACCCGTGGGGTGGAGTGGCCCTACTT 1380
Db 1321 AGGCTTTGGCATGGGCAAGTATACACAGCCCTGACCCGTGGGGTGGAGTGGCCCTACTT 1380
QY 1381 GGCCACTACGTGAGCTGGCACTTCCTTTTGGAGTCCAGAGGCCCAAGACAAATACGTGA 1440
Db 1381 GGCCACTACGTGAGCTGGCACTTCCTTTTGGAGTCCAGAGGCCCAAGACAAATACGTGA 1440
QY 1441 TGCCCTTTTGTGTTTGAACAGAAACAGAGGCTTCCCTGCGGCGCAACCACTCAGATCT 1500
Db 1441 TGCCCTTTTGTGTTTGAACAGAAACAGAGGCTTCCCTGCGGCGCAACCACTCAGATCT 1500
QY 1501 CTACTGCACTACTTTGGGGGCTTGGGAAAGGCTGTGCTGATCATATCTATCCAC 1560
Db 1501 CTACTGCACTACTTTGGGGGCTTGGGAAAGGCTGTGCTGATCATATCTATCCAC 1560
QY 1561 CTGTCTCAACTATGACTATGTGTGGATACGGTCTTCCACCCCACTGGGGCCATGAAAT 1620
Db 1561 CTGTCTCAACTATGACTATGTGTGGATACGGTCTTCCACCCCACTGGGGCCATGAAAT 1620
QY 1621 ACGATTCTATGCGACAGGGCTATCATAGCTCGGCAATTCCTTTGGTGTACTGGGAAATA 1680
Db 1621 ACGATTCTATGCGACAGGGCTATCATAGCTCGGCAATTCCTTTGGTGTACTGGGAAATA 1680
QY 1681 CGGGAACCAAGTGTAGAGACACACCTGGGCGAGCTCCACACCCAGCGCCCAATTGAA 1740
Db 1681 CGGGAACCAAGTGTAGAGACACACCTGGGCGAGCTCCACACCCAGCGCCCAATTGAA 1740
QY 1741 GGTGATCTGTGATGTAGCAGACTGAGAACTGGGCTTGGGCCGAGATATGTCTTGT 1800
Db 1741 GGTGATCTGTGATGTAGCAGACTGAGAACTGGGCTTGGGCCGAGATATGTCTTGT 1800
QY 1801 CCCCATGGCTGTGCCCTTGAAGCCCTGAGCAGCTGACAGAGGCTGCAAGTATACCCGGAA 1860
Db 1801 CCCCATGGCTGTGCCCTTGAAGCCCTGAGCAGCTGACAGAGGCTGCAAGTATACCCGGAA 1860
QY 1861 GCTGTGAGATGAGAGAGCAGAGCCGCTTCTCTGTGGGAAGCGCCACCCCTGCTACTCT 1920
Db 1861 GCTGTGAGATGAGAGAGCAGAGCCGCTTCTCTGTGGGAAGCGCCACCCCTGCTACTCT 1920
QY 1921 GTACTGTGCGCAACACAGCAACAGTGGGTACCCCGGGGCTACCGCATCAGAT 1980
Db 1921 GTACTGTGCGCAACACAGCAACAGTGGGTACCCCGGGGCTACCGCATCAGAT 1980

```

QY 1981 GCTCAGCTTTGCTGGAGAGCCGCTGCCCCAAAACAGCTCCATGGCGAGAGGCTTCACTG 2040
| | | | |
Db 1981 GCTCAGCTTTGCTGGAGAGCCGCTGCCCCAAAACAGCTCCATGGCGAGAGGCTTCACTG 2040
QY 2041 GGAGAGGTACCAAGCTGGCTGTGACCCAGGGAAGGAGGAGGCCAGTAGAGCAGCGCT 2100
| | | | |
Db 2041 GGAGAGGTACCAAGCTGGCTGTGACCCAGGGAAGGAGGAGGCCAGTAGAGCAGCGCT 2100
QY 2101 TTTCATATGAGTGAACCTTTGGGCCCCCACTGTGATTCAGTACGATTCATCAACATGA 2160
| | | | |
Db 2101 TTTCATATGAGTGAACCTTTGGGCCCCCACTGTGATTCAGTACGATTCATCAACATGA 2160
QY 2161 GACCAATGCTGGAAGAGATTTGGGCGCTGGGTGACAGCTGTTTCTGCATATCCACA 2220
| | | | |
Db 2161 GACCAATGCTGGAAGAGATTTGGGCGCTGGGTGACAGCTGTTTCTGCATATCCACA 2220
QY 2221 TGCAGAGGACATTCCTTAACAGATGACTGTGGGGAACGGCGTGGCTTCTCTCCGACC 2280
| | | | |
Db 2221 TGCAGAGGACATTCCTTAACAGATGACTGTGGGGAACGGCGTGGCTTCTCTCCGACC 2280
QY 2281 CTATAACTCTTTGAGAGAACCCCTCTTCTACTGTGCGACTCCATCTACTTCCGAG 2340
| | | | |
Db 2281 CTATAACTCTTTGAGAGAACCCCTCTTCTACTGTGCGACTCCATCTACTTCCGAG 2340
QY 2341 GGAACGAGATGCTGGGCGCTGCGAGGTCAACCCCTAGCTTGCTGCCGCCAGGCTGCTG 2400
| | | | |
Db 2341 GGAACGAGATGCTGGGCGCTGCGAGGTCAACCCCTAGCTTGCTGCCGCCAGGCTGCTG 2400
QY 2401 CTGTGCCCCGACCTCCCTGCGCTTCTCCACGGGGGCTTCTCAACATGAGCGGCT 2460
| | | | |
Db 2401 CTGTGCCCCGACCTCCCTGCGCTTCTCCACGGGGGCTTCTCAACATGAGCGGCT 2460
QY 2461 GGGATGGGAGATGTGGCCCAAGGGCTCCAGGGCCAGGGGTGTGAGGATGGGAGCAGCTG 2520
| | | | |
Db 2461 GGGATGGGAGATGTGGCCCAAGGGCTCCAGGGCCAGGGGTGTGAGGATGGGAGCAGCTG 2520
QY 2521 GCACTGGGCGGAGGCTGTTCCCTTCTTCTCTGTGCGAGACTCTTCTTCTCCACTAC 2580
| | | | |
Db 2521 GCACTGGGCGGAGGCTGTTCCCTTCTTCTCTGTGCGAGACTCTTCTTCTCCACTAC 2580
QY 2581 CCTCCCTGCATCCGCTCTGAGCCAGAGACCTCTGACCCCTGTGATGCTTGACACAGG 2640
| | | | |
Db 2581 CCTCCCTGCATCCGCTCTGAGCCAGAGACCTCTGACCCCTGTGATGCTTGACACAGG 2640
QY 2641 GACACTGAACCTTGTGATGCCAGCTGTACTGATGTTCTCATCCACAGAGGCCAGGATG 2700
| | | | |
Db 2641 GACACTGAACCTTGTGATGCCAGCTGTACTGATGTTCTCATCCACAGAGGCCAGGATG 2700
QY 2701 CCCAGGCTGAGGCGGTGGCGGAGGCTTCCCTAGATGTTCCCTTGTCTGTCTGCT 2760
| | | | |
Db 2701 CCCAGGCTGAGGCGGTGGCGGAGGCTTCCCTAGATGTTCCCTTGTCTGTCTGCT 2760
QY 2761 TTCCCAATCTTTTAAAGGCACTCCAAAGAGCTTAAAGGGGGCTTTCCCTGGAGACC 2820
| | | | |
Db 2761 TTCCCAATCTTTTAAAGGCACTCCAAAGAGCTTAAAGGGGGCTTTCCCTGGAGACC 2820
QY 2821 CCAGATAGAGGTGCCAGTCTCAGATGCTCATAGCTAGCTGGAAGAGATGTTCTCTC 2880
| | | | |
Db 2821 CCAGATAGAGGTGCCAGTCTCAGATGCTCATAGCTAGCTGGAAGAGATGTTCTCTC 2880
QY 2881 ACATTCCTCTCATCCAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
| | | | |
Db 2881 ACATTCCTCTCATCCAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
QY 2941 TCT 3000
| | | | |
Db 2941 TCT 3000
QY 3001 ATCCCTATGTCCAGCCCTGTGACTCCCGAGCCCTCAATTTTCACTCAAGTTCCGCT 3060
| | | | |
Db 3001 ATCCCTATGTCCAGCCCTGTGACTCCCGAGCCCTCAATTTTCACTCAAGTTCCGCT 3060
QY 3061 CCTCTCAGCCCTATGGAATCTCAAGTCAAGGAGCCCTTAATCAAGTGGCCATCC 3120

Db 3061 CCTCTCAGCCCTATGGAATCTCAAGTCAAGGAGCCCTTAATCAAGTGGCCATCC 3120
| | | | |
QY 3121 TGTGTGCTGCTTCCCTGTGTGCTGTGCTTATTTGGAGTAGAGTGTCTTACCCCTGTC 3180
| | | | |
Db 3121 TGTGTGCTGCTTCCCTGTGTGCTGTGCTTATTTGGAGTAGAGTGTCTTACCCCTGTC 3180
QY 3181 CTGGGCTGGGTGTGTTTCAGACAGCTGCTTCTGTGCATTTGTGTCGTGCTCCATG 3240
| | | | |
Db 3181 CTGGGCTGGGTGTGTTTCAGACAGCTGCTTCTGTGCATTTGTGTCGTGCTCCATG 3240
QY 3241 CTCTCTATAGAGAGAGTGTCTCTGTGACAGACAGCTCAAGTTACATTTCAAGTGA 3300
| | | | |
Db 3241 CTCTCTATAGAGAGAGTGTCTCTGTGACAGACAGCTCAAGTTACATTTCAAGTGA 3300
QY 3301 TTTGGGGGTGCATGATTAATGAAGATGGCCATTTGTATACAGGGCTGTATTTCCAA 3360
| | | | |
Db 3301 TTTGGGGGTGCATGATTAATGAAGATGGCCATTTGTATACAGGGCTGTATTTCCAA 3360
QY 3361 CAGCTGTGTTGGAGGCTGAGTGAACCAAGGGTGGCATCAAGATGAGAAGCCAAA 3420
| | | | |
Db 3361 CAGCTGTGTTGGAGGCTGAGTGAACCAAGGGTGGCATCAAGATGAGAAGCCAAA 3420
QY 3421 GCGCCTACAACTCCAGCCACCCAGCAGGAGGGGCTCTCAATTCATTTAGGATGCGA 3480
| | | | |
Db 3421 GCGCCTACAACTCCAGCCACCCAGCAGGAGGGGCTCTCAATTCATTTAGGATGCGA 3480
QY 3481 ATGAGCTGGGCTCTGGGTGAGTGGGGGTGTGGGCTGTGAGGAGGGGCTGCGCTGGT 3540
| | | | |
Db 3481 ATGAGCTGGGCTCTGGGTGAGTGGGGGTGTGGGCTGTGAGGAGGGGCTGCGCTGGT 3540
QY 3541 GGGGAGGGGCTGGCTGTGTCAGGCTTGGGCTCCATTTCCATCATCTGTGCTCTCTG 3600
| | | | |
Db 3541 GGGGAGGGGCTGGCTGTGTCAGGCTTGGGCTCCATTTCCATCATCTGTGCTCTCTG 3600
QY 3601 AGGCTGAGTGGGGATGGGGAGCAAGAAATAGCAAGATGAGAAACAGAACTTT 3660
| | | | |
Db 3601 AGGCTGAGTGGGGATGGGGAGCAAGAAATAGCAAGATGAGAAACAGAACTTT 3660
QY 3661 TTCTCTAAAGGAGCTGTTAAATCAATTCATGATACAGCTTACAAATAGATGACAG 3720
| | | | |
Db 3661 TTCTCTAAAGGAGCTGTTAAATCAATTCATGATACAGCTTACAAATAGATGACAG 3720
QY 3721 CTAATAAATATGATGCTTTATATATCTTAATATGATTTAGTTAGTGAAGGCT 3780
| | | | |
Db 3721 CTAATAAATATGATGCTTTATATATCTTAATATGATTTAGTTAGTGAAGGCT 3780
QY 3781 AAGGCACAGAAATGTATAGCGGCACTTCCCATTTGTGTTCCAGAAAGATGAAATTA 3840
| | | | |
Db 3781 AAGGCACAGAAATGTATAGCGGCACTTCCCATTTGTGTTCCAGAAAGATGAAATTA 3840
QY 3841 AACCATATATGCTTATATGATGCTTATACATATGATGATGATGATGATGATGATGAT 3900
| | | | |
Db 3841 AACCATATATGCTTATATGATGCTTATACATATGATGATGATGATGATGATGATGAT 3900
QY 3901 GAGGGGAACAGTAGTGTGAGGAGACAGAGAGGAGGAGGCTTTAACCTTACACCTTTTG 3960
| | | | |
Db 3901 GAGGGGAACAGTAGTGTGAGGAGACAGAGAGGAGGAGGCTTTAACCTTACACCTTTTG 3960
QY 3961 TACATTTGAATTTGAACATGACATGACATGATTTACATTTGAATTTGAATTTGA 4020
| | | | |
Db 3961 TACATTTGAATTTGAACATGACATGACATGATTTACATTTGAATTTGAATTTGA 4020
QY 4021 CCCAAAAAAAAAAAAAAAAAAAA 4040
| | | | |
Db 4021 CCCAAAAAAAAAAAAAAAAAAAA 4040

RESULT 2
AX333039 4040 bp DNA linear PAT 09-JAN-2002
LOCUS
DEFINITION Sequence 3548 from Patent WO0194629.
ACCESSION AX333039

VERSION AX33039.1 GI:18123673
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,
Horligan, S., Sopper, D. R., and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 3548 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source 1.4040
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 825 a 1183 c 1097 g 934 t 1 others
ORIGIN
Query Match 100.0%; Score 4039.6; DB 6; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 GTCCTCCACACCTTATGCCAGCATCTGACTACCGGGAACTCAGCAGACCGGGA 60
1 GTCTCTCCACCTTATGCCAGCATCTGACTACCGGGAACTCAGCAGACCGGGA 60
61 GCCCCCCACCCGTCAGAGCCAAAGAGCCCGCTTCTGCTGCGTGAATACATTG 120
61 GCCCCCCACCCGTCAGAGCCAAAGAGCCCGCTTCTGCTGCGTGAATACATTG 120
121 CTCCTCTTGTGTAATCAGCTGTCCCTCTTGTGGGAAAATGAACAGAGAACATCT 180
121 CTCCTCTTGTGTAATCAGCTGTCCCTCTTGTGGGAAAATGAACAGAGAACATCT 180
181 CGTCTCTCTATCTGTGGCGGTCATCACCATTTGCGCTGTGTGTGTCGAGTGGTGG 240
181 CGTCTCTCTATCTGTGGCGGTCATCACCATTTGCGCTGTGTGTGTCGAGTGGTGG 240
241 CAGGGGTGAGATGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAG 300
241 CAGGGGTGAGATGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAG 300
301 TGCCCCAGCTTGGACACACCTGGCCAGAGCAGCTGTTTGCAGACCTGAGCCAGAGA 360
301 TGCCCCAGCTTGGACACACCTGGCCAGAGCAGCTGTTTGCAGACCTGAGCCAGAGA 360
361 GCTACGAGCTGTGATGCGCTTTCGACCCAGCGGCTGGGGCCAGGGCTGTGTGATGACG 420
361 GCTACGAGCTGTGATGCGCTTTCGACCCAGCGGCTGGGGCCAGGGCTGTGTGATGACG 420
421 CCAGGCGCGGCGCTCGAGCAACTGTCTTCTCACTGAGTGGAGTTCAGCTCCCAAGGC 480
421 CCAGGCGCGGCGCTCGAGCAACTGTCTTCTCACTGAGTGGAGTTCAGCTCCCAAGGC 480
481 TGCAAGCCTGTGCTACTTGGACAGAGGGGAGGCCCCCACTTGGCGGGAGCACTGGCAT 540
481 TGCAAGCCTGTGCTACTTGGACAGAGGGGAGGCCCCCACTTGGCGGGAGCACTGGCAT 540
481 TGCAAGCCTGTGCTACTTGGACAGAGGGGAGGCCCCCACTTGGCGGGAGCACTGGCAT 540
541 CGTCTCTTGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
541 CGTCTCTTGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
601 TCACCCCTCTACATGCGGAGCTGACTGTGAGCGTATGAGAGGAGGAGGAGGAGGAGGAGG 660
601 TCACCCCTCTACATGCGGAGCTGACTGTGAGCGTATGAGAGGAGGAGGAGGAGGAGGAGG 660
661 CCGAGCGCGGCTGCTGCTTCCAGAGTACCTGAGATAGACAGATGATCTTCAACAGAGA 720
661 CCGAGCGCGGCTGCTGCTTCCAGAGTACCTGAGATAGACAGATGATCTTCAACAGAGA 720
721 GCTGCGCCAGGCTTGTGGGCTTCTCCACACAGCTGTGCTTCAACAGACCGGGAGCGGA 780

Db 721 GCTGCCCGACGCTTCTGGGCTTCTCCACACTGTGTCTTACAGACCGGGAGCGGA 780
781 CTTGGTACAAATGACACAGGCTCCCGTGTCTGCAATCAGGGAGCCGGCCACTGCTT 840
781 CTTGGTACAAATGACACAGGCTCCCGTGTCTGCAATCAGGGAGCCGGCCACTGCTT 840
841 TGGGCTTACTACAAATCTCGGGGCTGGGTCTTCTTCTCCACACTGGGGCTTGGAGCT 900
841 TGGGCTTACTACAAATCTCGGGGCTGGGTCTTCTTCTCCACACTGGGGCTTGGAGCT 900
901 GCTAGTAAACACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
901 GCTAGTAAACACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
961 AGCGCGCTACTACAGACAGCCTGCGCCAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
961 AGCGCGCTACTACAGACAGCCTGCGCCAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
1021 TGTGCTCTGATCCCAACAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
1021 TGTGCTCTGATCCCAACAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
1081 CCGCGGTCCAGCTCCCGCTCTACAGTCTATCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1081 CCGCGGTCCAGCTCCCGCTCTACAGTCTATCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1141 AAGTGAAGTGGGCTCTCTACTGTGAGCTTCTCTTGGGCTGGAGATTCAGTGGCCC 1200
1141 AAGTGAAGTGGGCTCTCTACTGTGAGCTTCTCTTGGGCTGGAGATTCAGTGGCCC 1200
1201 AAGGATCTTGTACCTTGGCTTCCCAAGAGAAAGCTAGTTATAGATTAAGCTTCCAAAGA 1260
1201 AAGGATCTTGTACCTTGGCTTCCCAAGAGAAAGCTAGTTATAGATTAAGCTTCCAAAGA 1260
1261 GGCCTTGGGCTCTATGATGAGGAAATCCCGAGCAATGAGAGGAGGAGGAGGAGGAGGAGGAG 1320
1261 GGCCTTGGGCTCTATGATGAGGAAATCCCGAGCAATGAGAGGAGGAGGAGGAGGAGGAGGAG 1320
1321 AGGCTTGGGCTCTATGATGAGGAAATCCCGAGCAATGAGAGGAGGAGGAGGAGGAGGAGGAG 1380
1321 AGGCTTGGGCTCTATGATGAGGAAATCCCGAGCAATGAGAGGAGGAGGAGGAGGAGGAGGAG 1380
1381 GGGCAGCTAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
1381 GGGCAGCTAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
1441 TGCCTTTTGTGTGTAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
1441 TGCCTTTTGTGTGTAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
1501 CTACTGCGACTACTTGGGGGCTTGGGAGAAACGCTGTGCTGCTCAGATCTATGCTCCAC 1560
1501 CTACTGCGACTACTTGGGGGCTTGGGAGAAACGCTGTGCTGCTCAGATCTATGCTCCAC 1560
1561 CTGCTCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
1561 CTGCTCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
1621 ACGATTCATATGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
1621 ACGATTCATATGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
1681 CCGGAGCAAAATGTGAGAGCAGACCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
1681 CCGGAGCAAAATGTGAGAGCAGACCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
1741 GGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
1741 GGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
1801 CCCCATGCTGTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
1801 CCCCATGCTGTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860


```
Db 1801 CCCCATGGCTGTGTCCTGGAGCCCTGAGCACACAGCTGCAGAGGCTGCAGGTGACCCGGAA 1860
Qy 1861 GCTGTGAGAGATGAGAGAGAGCCGCGCTTCTGTGGGAAGCGCACCCCTGCTACCT 1920
Db 1861 GCTGTGAGAGATGAGAGAGAGCCGCGCTTCTGTGGGAAGCGCACCCCTGCTACCT 1920
Qy 1921 GTACCTGGCCAGCAACACAGCAACAAGTGGGGTCAACCCCCGGGGGCTACCGCATCCAGAT 1980
Db 1921 GTACCTGGCCAGCAACACAGCAACAAGTGGGGTCAACCCCCGGGGGCTACCGCATCCAGAT 1980
Qy 1981 GCTCAGCTTGTGTGAGAGCCGCTGCCCCAAAACAGCTCCATGGCGAGAGCTTCACAGT 2040
Db 1981 GCTCAGCTTGTGTGAGAGCCGCTGCCCCAAAACAGCTCCATGGCGAGAGCTTCACAGT 2040
Qy 2041 GGAGAGGTACACAGCTGGCTGTGACCCAGCGGAAGGAGAGAGAGCCAGTAGACAGACAGCT 2100
Db 2041 GGAGAGGTACACAGCTGGCTGTGACCCAGCGGAAGGAGAGAGAGCCAGTAGACAGACAGCT 2100
Qy 2101 TTTCAATCAGAAATGACCTTGGGGCCCCAGTGTGATTTTCAATGATCTTCAATCAATGA 2160
Db 2101 TTTCAATCAGAAATGACCTTGGGGCCCCAGTGTGATTTTCAATGATCTTCAATCAATGA 2160
Qy 2161 GACCATGTGTGGAAGATTTGGTGGCTGGGTGACAGCTGTGTTTGTGATATCCACA 2220
Db 2161 GACCATGTGTGGAAGATTTGGTGGCTGGGTGACAGCTGTGTTTGTGATATCCACA 2220
Qy 2221 TGCAGAGACATTCCTAACAACAGTGTGTGGGGAACGGCGTGGCTTCTCTCGAAC 2280
Db 2221 TGCAGAGACATTCCTAACAACAGTGTGTGGGGAACGGCGTGGCTTCTCTCGAAC 2280
Qy 2281 CTATTAATCTTTTGAAGAGACCCCTCTTCTACTGTGCGAGACATCAATCACTCCGAG 2340
Db 2281 CTATTAATCTTTTGAAGAGACCCCTCTTCTACTGTGCGAGACATCAATCACTCCGAG 2340
Qy 2341 GGAACAGAGATGTGGGGCTGTGCGAGGTCAACCCCTAGCTTGCCTGCCCCAGCTGTGC 2400
Db 2341 GGAACAGAGATGTGGGGCTGTGCGAGGTCAACCCCTAGCTTGCCTGCCCCAGCTGTGC 2400
Qy 2401 CTGTGCCCCGACCTCCCTGCTTCTCCACGGGGCTTCTCTACAACATAGCGGTCT 2460
Db 2401 CTGTGCCCCGACCTCCCTGCTTCTCCACGGGGCTTCTCTACAACATAGCGGTCT 2460
Qy 2461 GGGATGGGGGATGTGGGCAAGGGGTCCAGGGGCAAGGGGTGAGGGGATGGGGAGAGCTGG 2520
Db 2461 GGGATGGGGGATGTGGGCAAGGGGTCCAGGGGCAAGGGGTGAGGGGATGGGGAGAGCTGG 2520
Qy 2521 GCACTGGGGCGGACGCTGTGTCTTCTTCTGTGCGAGAGCTCTTCTTCTCCACTAC 2580
Db 2521 GCACTGGGGCGGACGCTGTGTCTTCTTCTGTGCGAGAGCTCTTCTTCTCCACTAC 2580
Qy 2581 CCTCCCTGCGATCGGCTCTGAGCGCAGAGAGCTCTGTGACCTGTGATCCTGACACAGGG 2640
Db 2581 CCTCCCTGCGATCGGCTCTGAGCGCAGAGAGCTCTGTGACCTGTGATCCTGACACAGGG 2640
Qy 2641 GACACTGAACCTTGTATGTGACAGCTGTACTGAGTTCTCATTCACAGAGGCGCAGGCAATGG 2700
Db 2641 GACACTGAACCTTGTATGTGACAGCTGTACTGAGTTCTCATTCACAGAGGCGCAGGCAATGG 2700
Qy 2701 CCCAGCTGTAGAGCGGTGGCCAGAGGCTTCCCTAGATGGTTTCTTGTGTGTGTGTGTGT 2760
Db 2701 CCCAGCTGTAGAGCGGTGGCCAGAGGCTTCCCTAGATGGTTTCTTGTGTGTGTGTGTGT 2760
Qy 2761 TTCCCGAATCTTTTGAAGCCATTCAGAGACTTAAAGGGGCTATTTCCCTGAGAGCC 2820
Db 2761 TTCCCGAATCTTTTGAAGCCATTCAGAGACTTAAAGGGGCTATTTCCCTGAGAGCC 2820
Qy 2821 CCAAGATAGGTTTGGCAATCTCTGCAAGTCAATAGCTGAGCTGGAAGAGATGCTTCTGCTC 2880
Db 2821 CCAAGATAGGTTTGGCAATCTCTGCAAGTCAATAGCTGAGCTGGAAGAGATGCTTCTGCTC 2880
Qy 2881 ACATTTCCCTCATCCAGAGCTTCTTCTGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Db 2881 ACATTTCCCTCATCCAGAGCTTCTTCTGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Qy 2941 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3000
Db 2941 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3000
Qy 3001 ATCCCTATGTCCAGAGCCCTGTGACTCCCTCCAGCCCTCACTTTTCACTCAAGTTCGCTC 3060
Db 3001 ATCCCTATGTCCAGAGCCCTGTGACTCCCTCCAGCCCTCACTTTTCACTCAAGTTCGCTC 3060
Qy 3061 CCTCTCCAGCCCTATGAGAGTCTCAAGTCAAGGAGCCCTTAATCAGAGTGGCCATCC 3120
Db 3061 CCTCTCCAGCCCTATGAGAGTCTCAAGTCAAGGAGCCCTTAATCAGAGTGGCCATCC 3120
Qy 3121 TGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3180
Db 3121 TGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3180
Qy 3181 CTGGGGCTGGGTGTGTTTCAAGAGAGCTGTCTGTGCAATTTGTGTGCTGCTGCTCATG 3240
Db 3181 CTGGGGCTGGGTGTGTTTCAAGAGAGCTGTCTGTGCAATTTGTGTGCTGCTGCTCATG 3240
Qy 3241 CTCTCTATAGAGAGATGTGATCTGTGACAGCAGCAGCTCAAGTTCAGATTTCAAGTGA 3300
Db 3241 CTCTCTATAGAGAGATGTGATCTGTGACAGCAGCAGCTCAAGTTCAGATTTCAAGTGA 3300
Qy 3301 TTTGGGGGTGCAATGATTAATGAATGAGCATTTTGTACAGGGCTGTGATTCGCA 3360
Db 3301 TTTGGGGGTGCAATGATTAATGAATGAGCATTTTGTACAGGGCTGTGATTCGCA 3360
Qy 3361 CAGCCTGTTGGAGGCTGTGAGTGAAGCAAGGGGTGGGCAATCAAGATGGAAGCCAAA 3420
Db 3361 CAGCCTGTTGGAGGCTGTGAGTGAAGCAAGGGGTGGGCAATCAAGATGGAAGCCAAA 3420
Qy 3421 GCCCCCTACACTCCAGCCACCAGCAGAGAGGGCTGTCCATCAGATTCAGCATGGA 3480
Db 3421 GCCCCCTACACTCCAGCCACCAGCAGAGAGGGCTGTCCATCAGATTCAGCATGGA 3480
Qy 3481 ATGAGCTGGGGCTGTGGGAGAGTGGGGGTCTGGGCTAGTGGGGAGGGGGCTGGGGCT 3540
Db 3481 ATGAGCTGGGGCTGTGGGAGAGTGGGGGTCTGGGCTAGTGGGGAGGGGGCTGGGGCT 3540
Qy 3541 GGGGAGAGGCTGTGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCT 3600
Db 3541 GGGGAGAGGCTGTGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCT 3600
Qy 3601 AGGTCTGATTTGGGGATGGGAGCAAGAAATAGCAAGATGAGAAACAGAAACTTT 3660
Db 3601 AGGTCTGATTTGGGGATGGGAGCAAGAAATAGCAAGATGAGAAACAGAAACTTT 3660
Qy 3661 TTTCTCTAAGAGATGCTGTTAATCAATCTCTATACAGCTTCATCAATCAATAGTATGAG 3720
Db 3661 TTTCTCTAAGAGATGCTGTTAATCAATCTCTATACAGCTTCATCAATCAATAGTATGAG 3720
Qy 3721 CTAATAAATTAATGATGATCTTATATATATATATATATATATATATATATATATATATAT 3780
Db 3721 CTAATAAATTAATGATGATCTTATATATATATATATATATATATATATATATATATATAT 3780
Qy 3781 AAGCCACAGAAATGTATATAGCGCACTTCCATTTGTGTCTTCAGAAAGAGATAGATATA 3840
Db 3781 AAGCCACAGAAATGTATAGCGCACTTCCATTTGTGTCTTCAGAAAGAGATAGATATA 3840
Qy 3841 AACACATATATGCTTATATATATATATATATATATATATATATATATATATATATATAT 3900
Db 3841 AACACATATATGCTTATATATATATATATATATATATATATATATATATATATATATAT 3900
Qy 3901 GAGGGGAACAGTATGATGATGAGAGACAGAGAGAGAGAGGCTTAACTTACCTTACCTTTG 3960
Db 3901 GAGGGGAACAGTATGATGATGAGAGACAGAGAGAGAGAGGCTTAACTTACCTTACCTTTG 3960
Qy 3961 TACATTTTGAATTTTGAACCAATGTGACTGTATTTACCTTTTCAATAATAAATAAATGGG 4020
Db 3961 TACATTTTGAATTTTGAACCAATGTGACTGTATTTACCTTTTCAATAATAAATAAATGGG 4020
```

OY 4021 CCAAAAAAAAAAAAAA 4040
Db 4021 CCAAAAAAAAAAAAAA 4040

RESULT 3
LOCUS HSU39447 4040 bp mRNA linear PRI 08-JAN-1997
DEFINITION Human placenta copper monamine oxidase mRNA, complete cds.
ACCESSION U39447
VERSION U39447.1 GI:1399031
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 4040)
AUTHORS Zhang, X. and McIntire, W.S.
TITLE Cloning and sequencing of a copper-containing, topa
quinone-containing monamine oxidase from human placenta
JOURNAL Gene 179 (2), 279-286 (1996)
MEDLINE 97128319
PUBMED 8972912
REFERENCE 2 (bases 1 to 4040)
AUTHORS Zhang, X. and McIntire, W.
TITLE Direct Submission
SUBMITTED (25-OCT-1995) William McIntire, Molecular Biology Div.
(151-S), VA Medical Center, 4150 Clement Street, San Francisco, CA
94121, USA

FEATURES
source location/Qualifiers
1. 4040
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="placenta"
/clone_id="ClonTECH human placenta 5'-stretch plus cDNA
library in lambda-gt11, cat. no. HL3007b"
/note="mRNA isolated from placenta of a 25-year-old
caucasian mother"
161..2452
/codon_start=1
/product="copper monamine oxidase"
/protein_id="AAC50919.1"
/db_xref="GI:13990312"
/translation="MNOKTILVILILAVITFALVCVILVGRGGGSPSOLPHCPSPV
SPSNQPTHHGOSQLFADLSREETAVRFTQLRGGLVDAADARSDNCFVSELD
LPRRAALAHLDKSPPARALAVFFGRQPOVNSLVLVGLPHPSYMDVVERH
GGPLPYRRPVLFEQYLDIDQMFNRELPOASGLHRCCKFKHGRNLVYMTAPRL
QSGDRATWRFGLYNIISGAGFLHNGLELVNHRKALPARYTIOKVFYOGRYYSLAQ
LEQFEAGLVNVVLIIPDNGTGSWSLKSVPYGPAPPLQFYPQGRFVSQSVRSYASL
WTFSEGLGAFSPRIIPDNGRGERLVEISLOELATVGGSPAPAKMTTRYDGGPCK
KYPPLPRGDCPLATVYVDMHFLIESQARTITDACVFEQNGLEPLRHHSDLYSH
YFGSLATLVVRSMSLTLNLDVMDVTFHPSGAIERFATGYIISAFLEGATKYG
NVSSEHTLGVTHSAHKVDLVAGLENNWMAEDMVEFPAVWSEPHQRLQVTR
KLEMEBOAALFVGSATPRYLALASNSKNGHGRGYRIQLISFAGELPONSMAFG
FSMERYOLATORKKEEPPSSSVNONDPNAPTYDFSPINERTIAGELDVAWYTAGF
LHIDHARDINPTVVGNGVFLRPVNFDPEDPSFYASDSITFRGDDAGACEVNPILA
CLPOAACAPDLPRFSGHSN"

sig_peptide 161..217
misc_feature 1571..1573
/note="possible signal peptide"
/note="encodes Tyr residue that is converted to the topa
quinone cofactor"
4004..4009
4011..4016
polya_signal 825 a 1183 c 1097 g 934 t 1 others
BASE COUNT 825 a 1183 c 1097 g 934 t 1 others
ORIGIN

Query Match 100.0%; Score 4039.6; DB 9; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCTTCCACCTTACTGCCAGCATCTGACTACGGGAACCTCAGCCAGATGCCGGA 60
|||||

Db 1 GTCTTCCACCTTACTGCCAGCATCTGACTACGGGAACCTCAGCCAGATGCCGGA 60
OY 61 GCCCCCAACCCGTCACAGAGCCCAACAGACCCCGCTTCTGCGGTGAGAAATCATTTG 120
|||||
Db 61 GCCCCCAACCCGTCACAGAGCCCAACAGACCCCGCTTCTGCGGTGAGAAATCATTTG 120
OY 121 CTCTCTTGGTTGAATAGCTGTGCTCTTCTGCGGAAAATGAACAGAAACATCTT 180
|||||
Db 121 CTCTCTTGGTTGAATAGCTGTGCTCTTCTGCGGAAAATGAACAGAAACATCTT 180
OY 181 CGTGCTCTCATTTCTGGCGGTCATACCATCTTTGGTTGTGTCTGCTGTGTGG 240
|||||
Db 181 CGTGCTCTCATTTCTGGCGGTCATACCATCTTTGGTTGTGTCTGCTGTGTGG 240
OY 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTTCCCATTTGCCCTGTATCTCCAG 300
|||||
Db 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTTCCCATTTGCCCTGTATCTCCAG 300
OY 301 TGCCAGCTTGGACACACCTGCGCAAGCCAGCTGTTTGACAGCTGAGCCGAGAGA 360
|||||
Db 301 TGCCAGCTTGGACACACCTGCGCAAGCCAGCTGTTTGACAGCTGAGCCGAGAGA 360
OY 361 GCTGACGCTGATGCGCTTCTGACCCAGCGGCTGGGGCCAGGGGCTGTGATGCAGC 420
|||||
Db 361 GCTGACGCTGATGCGCTTCTGACCCAGCGGCTGGGGCCAGGGGCTGTGATGCAGC 420
OY 421 CCAGGCGCGGCGCTGCAACACTGTCTTCTCACTGAGTGTGAGCTGCTCCCAAGGC 480
|||||
Db 421 CCAGGCGCGGCGCTGCAACACTGTCTTCTCACTGAGTGTGAGCTGCTCCCAAGGC 480
OY 481 TGCAACCTTGCTCACTTGGACAGGGGGAGCCCGCCACTGCGGGAGGCACTGGCAT 540
|||||
Db 481 TGCAACCTTGCTCACTTGGACAGGGGGAGCCCGCCACTGCGGGAGGCACTGGCAT 540
OY 541 CGTCTCTTTGGCAAGGCAACCCCAAGCCCAAGTGTGAGTGTGGGGGCACTGCC 600
|||||
Db 541 CGTCTCTTTGGCAAGGCAACCCCAAGTGTGAGTGTGGGGGCACTGCC 600
OY 601 TCACCTCTTGGCAAGGCAACCCCAAGTGTGAGTGTGGGGGCACTGCC 660
|||||
Db 601 TCACCTCTTGGCAAGGCAACCCCAAGTGTGAGTGTGGGGGCACTGCC 660
OY 661 CCGAGCGCGGCTGTCGAAGAGTACTGTGACATACCAAGATGATCTTCAACAGGA 720
|||||
Db 661 CCGAGCGCGGCTGTCGAAGAGTACTGTGACATACCAAGATGATCTTCAACAGGA 720
OY 721 GCTGCGCGGCTTCTGCGGCTTCTCAGCACTGTGTTCTTACAAGCAACCGGGAGCGAA 780
|||||
Db 721 GCTGCGCGGCTTCTGCGGCTTCTCAGCACTGTGTTCTTACAAGCAACCGGGAGCGAA 780
OY 781 CCGTGTGACATGACACGCGCTCCCGGTGCTGTGCAATCAGGGGACCGGCGACCTGGTT 840
|||||
Db 781 CCGTGTGACATGACACGCGCTCCCGGTGCTGTGCAATCAGGGGACCGGCGACCTGGTT 840
OY 841 TGGCGCTACACAACTCTGGGGGCGTGGTCTCTTCTGACACACGCGGCTGGAGCT 900
|||||
Db 841 TGGCGCTACACAACTCTGGGGGCGTGGTCTCTTCTGACACACGCGGCTGGAGCT 900
OY 901 GCTACTGACCAAGAGCGCTTGAACCTGCGCGCTGAGCTATCCAGAAAGTGTCTATCA 960
|||||
Db 901 GCTACTGACCAAGAGCGCTTGAACCTGCGCGCTGAGCTATCCAGAAAGTGTCTATCA 960
OY 961 AGCGCGCTACTACGACAGCTGCGCCAGCTGAGAGGCCAGTTTGAAGCGCGCGTGGTAA 1020
|||||
Db 961 AGCGCGCTACTACGACAGCTGCGCCAGCTGAGAGGCCAGTTTGAAGCGCGCGTGGTAA 1020
OY 1021 TGTGTGTGATCCAGCAATGGCAGAGGTGGGCTGTGCTCCAGTATCCCGTGGTGC 1080
|||||
Db 1021 TGTGTGTGATCCAGCAATGGCAGAGGTGGGCTGTGCTCCAGTATCCCGTGGTGC 1080
OY 1081 CCGGGGTCACCTCCCGCTTCTACAGTTCTATCCCAAGCGCCCGCTTCACTGTCCAGG 1140
|||||
Db 1081 CCGGGGTCACCTCCCGCTTCTACAGTTCTATCCCAAGCGCCCGCTTCACTGTCCAGG 1140

Qy	1141	AAGTGAAGTGGGCTCTCTACTGTGACCTTTCCTTGGCCTTGAGAGCATTCAGTGGCC	1200
Db	1141	AAGTGAAGTGGGCTCTCTACTGTGACCTTTCCTTGGCCTTGAGAGCATTCAGTGGCC	1200
Qy	1201	AAGGATCTTGAAGCTTCGGTCCAAAGAGAAAGACTAGTTTATGATATAAGCCTCCAGA	1260
Db	1201	AAGGATCTTGAAGCTTCGGTCCAAAGAGAAAGACTAGTTTATGATATAAGCCTCCAGA	1260
Qy	1261	GGCCTTGGCCATCTATGTGTGAATTTCCCGACGACGATGACGACCCGCTATGTGATGG	1320
Db	1261	GGCCTTGGCCATCTATGTGTGAATTTCCCGACGACGATGACGACCCGCTATGTGATGG	1320
Qy	1321	AGGCTTTGGCATGGGCAATATACACAGCGCCCTGAGCCGGGGGGTGGACTGGCCCTACTT	1380
Db	1321	AGGCTTTGGCATGGGCAATATACACAGCGCCCTGAGCCGGGGGGTGGACTGGCCCTACTT	1380
Qy	1381	GGCCACCTTACGTGAGCTGGCACTTCTTTTGAAGTCCAGGCCCCCAAGCAATACGTGA	1440
Db	1381	GGCCACCTTACGTGAGCTGGCACTTCTTTTGAAGTCCAGGCCCCCAAGCAATACGTGA	1440
Qy	1441	TGGCTTTTGTGTGTTGAACAGAACAGAGGGCCCTCCCGTGGGGGAGACACCACTACGATCT	1500
Db	1441	TGGCTTTTGTGTGTTGAACAGAACAGAGGGCCCTCCCGTGGGGGAGACACCACTACGATCT	1500
Qy	1501	CTACTCGCACTACTTTGGGGGGTTTGCAGAAAGGATGCTGTGTCAGATCTATGTGCAC	1560
Db	1501	CTACTCGCACTACTTTGGGGGGTTTGCAGAAAGGATGCTGTGTCAGATCTATGTGCAC	1560
Qy	1561	CTTGCTCAACTATGACTATGTGTGGGATACGGTCTTCCACCCCAAGTGGGCGCATAGAAAT	1620
Db	1561	CTTGCTCAACTATGACTATGTGTGGGATACGGTCTTCCACCCCAAGTGGGCGCATAGAAAT	1620
Qy	1621	ACGATTTCTATGCCACGGGCTACATCAGCTCGGCACTTCTTTGTGTCTACTGGGAAGTA	1680
Db	1621	ACGATTTCTATGCCACGGGCTACATCAGCTCGGCACTTCTTTGTGTCTACTGGGAAGTA	1680
Qy	1681	CGGGAAACAAGTGTACAGACACACCTGGGCGACGGTCCACACCACAGCGCCCACTTCAA	1740
Db	1681	CGGGAAACAAGTGTACAGACACACCTGGGCGACGGTCCACACCACAGCGCCCACTTCAA	1740
Qy	1741	GGGGATCTGTGGATGTAGACAGAGACTGGAGAACTGGGCTGGGGCGAGGATATAGCTTTGCT	1800
Db	1741	GGGGATCTGTGGATGTAGACAGAGACTGGAGAACTGGGCTGGGGCGAGGATATAGCTTTGCT	1800
Qy	1801	CCCCATAGGGTGTCCCTGTGAGCCCTGAGACACAGCTGCAGAGGCTGCAGAGTGACCCGGAA	1860
Db	1801	CCCCATAGGGTGTCCCTGTGAGCCCTGAGACACAGCTGCAGAGGCTGCAGAGTGACCCGGAA	1860
Qy	1861	GCTGCTGTGAGATGTGAGAGAGCAGGCGCCTTCTGTGTGGAAAGGCGCACCCCTGCTACTT	1920
Db	1861	GCTGCTGTGAGATGTGAGAGAGCAGGCGCCTTCTGTGTGGAAAGGCGCACCCCTGCTACTT	1920
Qy	1921	GTACCTGGCCAGCAACACAGCAAAAGTGGGGTACCCCCGGGGGCTTACCGCATTCAGAT	1980
Db	1921	GTACCTGGCCAGCAACACAGCAAAAGTGGGGTACCCCCGGGGGCTTACCGCATTCAGAT	1980
Qy	1981	GCTCAGCTTTGTGTGAGAGCGCGTGCCCCAATAAGCTCATGGCGAGAGGCTTCAAGTGG	2040
Db	1981	GCTCAGCTTTGTGTGAGAGCGCGTGCCCCAATAAGCTCATGGCGAGAGGCTTCAAGTGG	2040
Qy	2041	GGAGAGGTACCACTGGCTGTGACCCACGCGGAAGGAGGAGGAGGCCAGTACGACAGGCT	2100
Db	2041	GGAGAGGTACCACTGGCTGTGACCCACGCGGAAGGAGGAGGAGGCCAGTACGACAGGCT	2100
Qy	2101	TTTCAATCAGAATGACCTTTGGGGCCCCACTGTGTGATTTAGTACTTTCATCAACAATGA	2160
Db	2101	TTTCAATCAGAATGACCTTTGGGGCCCCACTGTGTGATTTAGTACTTTCATCAACAATGA	2160
Qy	2161	GACCATTTGCTGGAAGAAGATTTTGGTGGCTGGGTGACAGCTGTGTTTCTGCATATCCACA	2220
Db	2161	GACCATTTGCTGGAAGAAGATTTTGGTGGCTGGGTGACAGCTGTGTTTCTGCATATCCACA	2220

QY	2221	TCGAGAGCAATTCCTAAACACAGTACGTGTGGGGAACGGCGCTGGGCTTCTCTCCGACC	2280
Db	2221	TCGAGAGCAATTCCTAAACACAGTACGTGTGGGGAACGGCGCTGGGCTTCTCTCCGACC	2280
QY	2281	CTATACTTTTGGAGGAAGACCCCTCTTACCTGCGACATCCATCTACTCCGAGG	2340
Db	2281	CTATACTTTTGGAGGAAGACCCCTCTTACTCTCGCACTCCATCTACTCCGAGG	2340
QY	2341	GGACCAAGATGCTGGGGGCTGGCGAGGTCAACCCCTTACCTTGCCGCCCCAGGCTGTGC	2400
Db	2341	GGACCAAGATGCTGGGGGCTGGCGAGGTCAACCCCTTACTTGCTGCCCCAGGCTGTGC	2400
QY	2401	CTGTGCCCCGACCTCCCTGCTCTCTCCAGGGGGCTTCTCTCAACAATAAGCGGTCT	2460
Db	2401	CTGTGCCCCGACCTCCCTGCTCTCTCCAGGGGGCTTCTCTCAACAATAGGCGGTCT	2460
QY	2461	GGGATGGGGGATGTGGGCGAAAGGGCTCAGAGGGCAGGGGTGTAGGGATGGGAGAGCTGG	2520
Db	2461	GGGATGGGGGATGTGGGCGAAAGGGCTCAGAGGGCAGGGGTGTAGGGATGGGAGAGCTGG	2520
QY	2521	GCACGTGGGCGGAGCCTGGTGTCCCTTCTCTGTGCGAGGACTCTCTTCTCCACTAC	2580
Db	2521	GCACGTGGGCGGAGCCTGGTGTCCCTTCTCTGTGCGAGGACTCTCTTCTCCACTAC	2580
QY	2581	CCCTCCCTTCGATCCGCTCTGAGCCAGAGACCTCCTGACCTGTGATCCCTGACACAGG	2640
Db	2581	CCCTCCCTTCGATCCGCTCTGAGCCAGAGACCTCCTGACCTGTGATCCCTGACACAGG	2640
QY	2641	GACACTAACCTTGTATATCCAGCGTACGTACGTATCTCATCCACAGAGGCGACGACATGG	2700
Db	2641	GACACTAACCTTGTATATCCAGCGTACGTACGTATCTCATCCACAGAGGCGACGACATGG	2700
QY	2701	CCCAAGCTGGAGCGGTGGCGAGGGCTTCCCTAGATAGTGTCCCTTGTGTGCTGTGGCT	2760
Db	2701	CCCAAGCTGGAGCGGTGGCGAGGGCTTCCCTAGATAGTGTCCCTTGTGTGCTGTGGCT	2760
QY	2761	TTCCCGAATCTTTTATAGCCACCTCCAGAGACTTAAAGGGGCTATTCCTGTGAGAC	2820
Db	2761	TTCCCGAATCTTTTATAGCCACCTCCAGAGACTTAAAGGGGCTATTCCTGTGAGAC	2820
QY	2821	CCAAGATAGGTTGCCAGTCCCTGCAAGTCCATAGCTGAGCTGGAAAGGATCGTCTGCTC	2880
Db	2821	CCAAGATAGGTTGCCAGTCCCTGCAAGTCCATAGCTGAGCTGGAAAGGATCGTCTGCTC	2880
QY	2881	ACATTCCTCTCATCCAGAGCTCTTCTCTGTGCTTCTCTCTCATCACTACTCTCC	2940
Db	2881	ACATTCCTCTCATCCAGAGCTCTTCTCTGTGCTTCTCTCTCTCATCACTACTCTCC	2940
QY	2941	TCCTCTCTCTGTTCTGCTCTCTCTCTATCTGCAATTTCTCCGAATCCTGAGGGAT	3000
Db	2941	TCCTCTCTCTGTTCTGCTCTCTCTCTCTATCTGCAATTTCTCCGAATCCTGAGGGAT	3000
QY	3001	ATCCCTATGTCCCAAGCCCCCTGTGTACTGCCCAAGCTTTCAGTCAAGTTCGCCGT	3060
Db	3001	ATCCCTATGTCCCAAGCCCCCTGTGTACTGCCCAAGCTTTCAGTCAAGTTCGCCGT	3060
QY	3061	CCTCTCCAGCCCTATGGAAGTCTCAAGAGTCAAGGGACCCCTATCAGAGTGGCCAATCC	3120
Db	3061	CCTCTCCAGCCCTATGGAAGTCTCAAGAGTCAAGGGACCCCTATCAGAGTGGCCAATCC	3120
QY	3121	TGTGTGTGTTCCCTTGTGTCTTGTGCTTATTTGGAGTATGGAGTGTCTCTACCCCTGTG	3180
Db	3121	TGTGTGTGTTCCCTTGTGTCTTGTGCTTATTTGGAGTATGGAGTGTCTCTACCCCTGTG	3180
QY	3181	CTGGGGGCTGGGTGTCTTCAAGACAGCTGTTGTGTGCAATTTGTGTCTGCGCCTCATG	3240
Db	3181	CTGGGGGCTGGGTGTGTCTTCAAGAGAGCTGCTTGTGTGCAATTTGTGTCTGCGCCTCATG	3240
QY	3241	CTCTCTATAGAGGAGATGGTCAATCGGACAGAGAGGTCAAAGTATGACATTTCAAGTGA	3300
Db	3241	CTCTCTATAGAGGAGATGGTCAATCGTACAGAGAGGTCAAAGTATGACATTTCAAGTGA	3300
QY	3301	TTTGGGGGTGCANAATATGAGAAATGGCAATTTTGTATCCAGGGCTCTGTATTTCTGCA	3360

[illegible]

JOURNAL Submitted (20-MAY-1998) Biotite Therapies Ltd, Tylistokatu 6, Turku
FIN-20520, Finland

FEATURES

source Location/Qualifiers

1..2501

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q21"
/tissue_type="lung"
1..2501

/gene="VAP1"
/gene="VAP1"
81..2372

/EC_number="1.4.3.6"
/note="semicarbazide sensitive amine oxidase;
copper-containing amine oxidase homolog"
/codon_start=1
/product="vascular adhesion protein-1"
/protein_id="AAC25170.1"
/db_xref="GI:3283370"

/translation="MNKTKTLLIIVITFALVYLYVGRGDGGEPSQLPHCSV
SPSQPMTTHRQSQLPFDLSREELTANRFLTQRLGGLVDAQARSDNCVSELD
LIPPAALAHLDRESPPARPALAVFFGQPOPNVELVGLPHPSYMRDVTVERH
GGPLPHRRPLVLEQSEYLDIDQMLENRELQASGLHMCFFKHNRLVMTAPRL
QSGDRATVFGLYYINISGAGFELHNVGELLVNHKALDPAWMTIQKYVQGRYVLSLQ
LEAFEGALVNVVLIPIDNGTGSWSLSPVPGCAPLQRYPGQRFSGRYASLFL
WTFSEFGATSGEPRIFDVRQGERLVEISLEELAITYGNSPAAMTRVYDGFPMG
KMTPLRGVDCPYLATVYDMHFLLESQAKRTIDACVCEQNGDPLRHHSLSYH
YFGLAETVLYVRMSSTLLINDIYMDTVPHRSGAIEIRPATGITSALFPGATGTG
NOVSEHTLQVHTHSAHFKVDLDVAGLENNWMAEDVFPVAIVWSEHDLQLOVTR
KLEMEQDAFLVGSATPRYLLIYLSNHSNMWGRGRYRIQLSPAGPLPQNSMARG
FSIMRYOLIAVORKEEPSSTSYFNQNDPMAPTVDFSDFTINNETIACKDLVAMTAAQF
LHMRIADPMTVGVNGVGFELRPYNFPEDEDPSTFYSDSIFYRGDDADACEVNPILA
CLPQAAACAPDLPLPFSHGFSHN"

BASE COUNT 464 a 782 c 719 g 536 t

ORIGIN

Query Match 61.9%; Score 2501; DB 9; Length 2501;
Best Local Similarity 100.0%; Pired. No. 0;
Matches 2501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 81 GCCACAGAGCCCCCGCTGTCGTGGCGTGAAGATTCATTGCTCCTTGGTTGAATAG 140
|||||

1 GCCAACAGAGCCCCCGCTTGTGCTGGCGTGAAGATTCATTGCTCCTTGGTTGAATAG 60
|||||

141 CTGTCCCTCTTCTGTTGGAAAAAGAACCAAGAACAACTCTGTGCTCCTCATTTGCGCG 200
|||||

61 CTGTCCCTCTTCTGTTGGAAAAAGAACCAAGAACAACTCTGTGCTCCTCATTTGCGCG 120
|||||

Db 201 TCATACCATCTTTTCCCTTGGTTTGTGTCTGCTGCTGTCGGCAGAGGGGTGAGATGGGGGTG 260
|||||

121 TCATACCATCTTTTCCCTTGGTTTGTGTCTGCTGCTGTCGGCAGAGGGGTGAGATGGGGGTG 180
|||||

261 AACCCAGCAGGTTCCCATTTGCCCTCTGATATCTCCAGTGCAGCCAGCTTGGAGACAC 320
|||||

181 AACCCAGCAGCTTCCCATTTGCCCTCTGATATCTCCAGTGCAGCCAGCTTGGAGACAC 240
|||||

321 CTGGCCAGAGCCAGCTGTTTGCAGACTGAGCCGAGAGAGAGACTGAAGCTGTGATGCGCT 380
|||||

Db 241 CTGGCCAGAGCCAGCTGTTTGCAGACTGAGCCGAGAGAGAGAGCTGAGTGAAGCGCT 300
|||||

381 TTCTACACCCAGAGGGGTGGGGCCAGGGGCTGCTGATGACGCCAGGCCCGGGCCCTTGAC 440
|||||

Db 301 TTCTACACCCAGAGGGGTGGGGCCAGGGGCTGCTGATGACGCCAGGCCCGGGCCCTTGAC 360
|||||

441 ACTGTGTCCTTCATCTGAGAGTTGACAGTGCCTCCCAAGGCTGACAGCCCTGTGCTACTTGG 500
|||||

Db 361 ACTGTGTCCTTCATCTGAGAGTTGACAGTGCCTCCCAAGGCTGACAGCCCTGTGCTACTTGG 420
|||||

501 ACAGGGGAGAGCCCCCACCCTGCGCGGAGAGGACATGGCCATGCTTTTGGACAGGCAAC 560
|||||

Db 421 ACAGGGGAGAGCCCCCACCCTGCGCGGAGAGGACATGGCCATGCTTTTGGACAGGCAAC 480
|||||

ORGANISM clone:HEP11945.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2538)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Source 1..2538
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP11945"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/note="Cloning vector pME18FL13"
1..2538
misc_feature
/note="Highly similar to HSU39447 Human placenta copper
monamine oxidase mRNA"
BASE COUNT 554 a 687 c 696 g 601 t
ORIGIN
Query Match 56.4%; Score 2277; DB 9; Length 2538;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2280; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1756 AGCAGAGCTGAGAACTGGGTGTGGCCGAGAGATAGTCTTTGCCCATGGCTGGCC 1815
DB 246 AGCAGAGCTGAGAACTGGGTGTGGCCGAGAGATAGTCTTTGCCCATGGCTGGCC 305
QY 1816 CTGAGGCGCTGAGCAAGCTCCAGAGCTGAGGTGACCCGGAAGCTGCTGAGATGGA 1875
DB 306 CTGAGGCGCTGAGCAAGCTCCAGAGCTGAGGTGACCCGGAAGCTGCTGAGATGGA 365
QY 1876 GGAGCAGGCGCGCTTCCTGCTGGGAAGCGCCACCCTGCTACCTGTAACCTGGCCAGCA 1935
DB 366 GGAGCAGGCGCGCTTCCTGCTGGGAAGCGCCACCCTGCTACCTGTAACCTGGCCAGCA 425
QY 1936 CCACAGCAACAAGTGGGTACACCCCGGGGCTACCGCATCCAGATGCTCAGCTTTGCTGG 1995
DB 426 CCACAGCAACAAGTGGGTACACCCCGGGGCTACCGCATCCAGATGCTCAGCTTTGCTGG 485
QY 1996 AGAGCGGCTGCCCCAAAACAGTCCATGGCGAGAGCTTCACCTGGGAGGTACCAAGT 2055
DB 486 AGAGCGGCTGCCCCAAAACAGTCCATGGCGAGAGCTTCACCTGGGAGGTACCAAGT 545
QY 2056 GCGCTGACCCAGCGGAAGAGAGAGAGCCAGTAGCAGCAGCTTTTCAATCAGAAATGA 2115
DB 546 GCGCTGACCCAGCGGAAGAGAGAGAGCCAGTAGCAGCAGCTTTTCAATCAGAAATGA 605
QY 2116 CCCTTGGGCCCCCAGCTGTGATTTTCAGTCTTCAACAATGAGACCATTTGCTGAAA 2175
DB 606 CCCTTGGGCCCCCAGCTGTGATTTTCAGTCTTCAACAATGAGACCATTTGCTGAAA 665

QY 2176 GCATTTGGTGGCGCTGGGTGACAGCTGGTTTTTCTGCATATCCCATGAGAGACATTC 2235
DB 666 GCATTTGGTGGCGCTGGGTGACAGCTGGTTTTTCTGCATATCCCATGAGAGACATTC 725
QY 2236 TAACACAGGACTGGGGGAAGGGGTGGGCTTCTCCGACCCCTAATCTTTGA 2295
DB 726 TAACACAGGACTGGGGGAAGGGGTGGGCTTCTCCGACCCCTAATCTTTGA 785
QY 2296 CGAAGACCCCTCTCTACTCTGCGACATCATCTACTTCCGAGGGAGCAGAGATGTGG 2355
DB 786 CGAAGACCCCTCTCTACTCTGCGACATCATCTACTTCCGAGGGAGCAGAGATGTGG 845
QY 2356 GCGCTGCGAGTCAACCCCTTACGTTGCTGCCCCAGCGCTGCTGCTGCCCCGACCT 2415
DB 846 GCGCTGCGAGTCAACCCCTTACGTTGCTGCCCCAGCGCTGCTGCTGCCCCGACCT 905
QY 2416 CCTGCGTTCTCCACGCGGGGCTTCTCACAACATAGGGGTCGAGGATGGGGCATG 2475
DB 906 CCTGCGTTCTCCACGCGGGGCTTCTCACAACATAGGGGTCGAGGATGGGGCATG 965
QY 2476 GCCAAGGCTCCAGGGCCAGAGGTGTGAGGAGATGGGAGCAGCTGGGCACTGGCGCGAG 2535
DB 966 GCCAAGGCTCCAGGGCCAGAGGTGTGAGGAGATGGGAGCAGCTGGGCACTGGCGCGAG 1025
QY 2536 CCTGCTCCCTTCTCTCTGTCGACAGACTCTTTTCTTCCACTACCTCCCTGCAATCG 2595
DB 1026 CCTGCTCCCTTCTCTCTGTCGACAGACTCTTTTCTTCCACTACCTCCCTGCAATCG 1085
QY 2596 CCTGTAGCAGAGACCTCTGACCTGATGATGCTGACACAGGGGACATGAACCTGT 2655
DB 1086 CCTGTAGCAGAGACCTCTGACCTGATGATGCTGACACAGGGGACATGAACCTGT 1145
QY 2656 TGATGCGAGCTGACTGAGTTCTATCCACAGAGAGCCAGGACAGCTGGAGCG 2715
DB 1146 TGATGCGAGCTGACTGAGTTCTATCCACAGAGAGCCAGGACATGGCCAGCTGGAGCG 1205
QY 2716 TGGCGGAGGCTTCCCTAGATAGTTCCTTTTGTCTGTGCTGCTTCCCAATCTTTT 2775
DB 1206 TGGCGGAGGCTTCCCTAGATAGTTCCTTTTGTCTGTGCTGCTTCCCAATCTTTT 1265
QY 2776 AGGCCACCTCCAAAGACCTTAAAGGGGCTATTCCTGAGAGACCCAGAGTAGGGTTGC 2835
DB 1266 AGGCCACCTCCAAAGACCTTAAAGGGGCTATTCCTGAGAGACCCAGAGTAGGGTTGC 1325
QY 2836 CAGTCTGCAAGTCAATAGCTGAGCTGGAAGAGATGCTTCTGCAATTCCTCTATC 2895
DB 1326 CAGTCTGCAAGTCAATAGCTGAGCTGGAAGAGATGCTTCTGCAATTCCTCTATC 1385
QY 2896 CAGTCTTCTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2955
DB 1386 CAGTCTTCTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
QY 2956 TGCCTTCTCTTATCTGCAATTTCTCCGAATCTGAGGGGATATCCATATGCCAG 3015
DB 1446 TGCCTTCTCTTATCTGCAATTTCTCCGAATCTGAGGGGATATCCATATGCCAG 1505
QY 3016 CCCCTGTACTCCCCAGACCCCTCACTTTTCAAGTCAAGTCCGTCCTCTCCAGCCAT 3075
DB 1506 CCCCTGTACTCCCCAGACCCCTCACTTTTCAAGTCAAGTCCGTCCTCTCCAGCCAT 1565
QY 3076 GGAAGTCTCAAGGTACAGGAGCCCTTAATCAGAGTGGCCAAATCCCTGTGTCTCCCT 3135
DB 1566 GGAAGTCTCAAGGTACAGGAGCCCTTAATCAGAGTGGCCAAATCCCTGTGTCTCCCT 1625
QY 3136 TGTGTCTGTGCTTATTTGGAGTAGAATGCTCTTACCCCTGTCTGGGGCTGGGTG 3195
DB 1626 TGTGTCTGTGCTTATTTGGAGTAGAATGCTCTTACCCCTGTCTGGGGCTGGGTG 1685
QY 3196 TTTCAAGACAGCTGCTTGTGATTTTGTCTGCTGCTCTCATATGCTCTTATAGAGAG 3255
DB 1686 TTTCAAGACAGCTGCTTGTGATTTTGTCTGCTGCTCTCATATGCTCTTATAGAGAG 1745
QY 3256 GATGTCATCTGAGACGACGACTCAAGTTAGCATTTTCAAGTATTTGGGGGTGCAATG 3315

```

|||||
Db 1746 GATGTCATCTGACACAGACAGCTCAAGTACATTCAAGGATTTGGGGTGCATG 1805
|||
Qy 3316 ATATAGAAATGGCCATTTTGTACAGGGCTGTATTCGTGCACAGCCCTGTTGGAG 3375
|||||
Db 1806 ATATAGAAATGGCCATTTTGTACAGGGCTGTATTCGTGCACAGCCCTGTTGGAG 1865
|||||
Qy 3376 GCTGGAGTGGAAACAAAGGTGGGCATCAAAAGTGAAGAACCCCAAGCCCTTACACTCCA 3435
|||||
Db 1866 GCTGGAGTGGAAACAAAGGTGGGCATCAAAAGTGAAGAACCCCAAGCCCTTACACTCCA 1925
|||||
Qy 3436 GCCACCCAGCAGAGGGGGTGTCAATCATTCATTCAGGCATGGCAATGAGCTGGCCCTG 3495
|||||
Db 1926 GCCACCCAGCAGAGGGGGTGTCAATCATTCATTCAGGCATGGCAATGAGCTGGCCCTG 1985
|||||
Qy 3496 GGTGAGGTGGGGTGTGGCTTGTGGAGGGGCTGGCTGGGTGGGGCAGGGCCCTGGC 3555
|||||
Db 1986 GGTGAGGTGGGGTGTGGCTTGTGGAGGGGCTGGCTGGGTGGGGCAGGGCCCTGGC 2045
|||||
Qy 3556 CTGGTCCAGGCTTGGGCTCCATTCCTCATCTGTCTCTCTGAGGCTGTGATTTGGG 3615
|||||
Db 2046 CTGGTCCAGGCTTGGGCTCCATTCCTCATCTGTCTCTCTGAGGCTGTGATTTGGG 2105
|||||
Qy 3616 ATGGGCAAAAGAAATAGCAAGAGATGAGAAACAGAAACTTTTCTCTAAAGACT 3675
|||||
Db 2106 ATGGGCAAAAGAAATAGCAAGAGATGAGAAACAGAAACTTTTCTCTAAAGACT 2165
|||||
Qy 3676 GGTAAATCAATTCGTATACAGCCTTACAATACATAGTATGACGCTAAATAATTTGT 3735
|||||
Db 2166 GGTAAATCAATTCGTATACAGCCTTACAATACATAGTATGACGCTAAATAATTTGT 2225
|||||
Qy 3736 ATCTCTTATATCTATATATATATATCTTCAGGTGTAAGGCAAGCCACAGAAATGT 3795
|||||
Db 2226 ATCTCTTATATCTATATATATATATCTTCAGGTGTAAGGCAAGCCACAGAAATGT 2285
|||||
Qy 3796 GTATAGGCACTTCCCTATTTGTGTTCAGAAAGAGTAGAATATTAACACATAATTTGCT 3855
|||||
Db 2286 GTATAGGCACTTCCCTATTTGTGTTCAGAAAGAGTAGAATATTAACACATAATTTGCT 2345
|||||
Qy 3856 ATGATGCTTATTCAGAAATTAATGAGTACAGTATTTGTTGGAGGGGAACCACTAG 3915
|||||
Db 2346 ATGATGCTTATTCAGAAATTAATGAGTACAGTATTTGTTGGAGGGGAACCACTAG 2405
|||||
Qy 3916 GTTGAAGCAGAGAGGGAAGGCTTAACACTTACACCTTTTGTACATTTGAATTTT 3975
|||||
Db 2406 GTTGAAGCAGAGAGGGAAGGCTTAACACTTACACCTTTTGTACATTTGAATTTT 2465
|||||
Qy 3976 GAACCATGTACTGTATTTACCTATTTCAAAATTAACAATAATGAGCCCAAAAAAAA 4035
|||||
Db 2466 GAACCATGTACTGTATTTACCTATTTCAAAATTAACAATAATGAGCCCAAAAAAAA 2525
|||||
Qy 4036 AAAAA 4040
|||||
Db 2526 AAAAA 2530

```

RESULT 6
 AB019242
 LOCUS
 DEFINITION
 Bos taurus SSAO mRNA for semicarbazide-sensitive amine oxidase,
 complete cds.
 AB019242
 ACCESSION
 AB019242.1 GI:6682824
 VERSION
 semicarbazide-sensitive amine oxidase: SSNO.
 KEYWORDS
 Bos taurus tissue; lib:Aorta CDNA to mRNA.
 SOURCE
 Bos taurus
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (sites)
 REFERENCE
 Iwabuchi, H., Matsumura, K., Mure, M., Kuroda, S. and Tanizawa, K.
 AUTHORS
 TITLE
 Molecular cloning of semicarbazide-sensitive amine oxidase gene
 from Bovine aorta

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3624)
 AUTHORS Iwabuchi, H., Matsumura, K., Kuroda, S. and Tanizawa, K.
 TITLE Direct Submission
 JOURNAL Submitted (27-Oct-1998) Shun-ichi Kuroda, Department of Structural
 Molecular Biology, Institute of Scientific and Industrial Research,
 Osaka Univ.; 8-1, Mihogaoka, Ibaraki, Osaka 567-0047, Japan
 (E-mail: skuroda@kenken.osaka-u.ac.jp, Tel.: +81-6-879-8461,
 Fax: +81-6-879-8464)

FEATURES
 Location/Qualifiers
 source
 1..3624
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /tissue_lib="Aorta"
 1..3624
 /gene="SSAO"
 1..3624
 /gene="SSAO"
 /note="Membrane-bound protein, Bovine Serum Amine Oxidase
 homolog, Vascular Adhesion Protein homolog, Quinoprotein"
 190..2481
 /gene="SSAO"
 /EC_number="1.4.3.6"
 /product="1"
 /codon_start=1
 /protein_id="BA88896.1"
 /translation="MNQKTLVLLALAVITITFALVCVLIAGRGDGEASOPHYCBG
 TPSVQWTHPQGNQLEPADLSRELTAVMSFLQKLEPDLVDAQAPSDNCIEVELQ
 LPPKAAALHLDRSPPARALAIYFEGGQOPNTVELVGPLQPSYMRDVTVERH
 GGLPYRRPVLRLREYLDIDOMFENELPOAAGVHLHCCSYKGGGKNTVMTAPGL
 OSGDRAFMELGYLNTISAGAYLAPVGLLELDLNDHKALDPAQMTLQKVEFGRTYSLQ
 LERQEFAGRVNRYVTPNNSTGSSMSKSVPPGPPPLDHPHGGRTFSVQGRYSLSL
 WTFSFELGAFSEPRIPDIRFQGRLELISLDAVAITGNTPAALNRYMGCTRMG
 KRATPLRGVDCPYLATYVDVHFLLESOAPRLDHFVCEQNKGLPLRHHSDFISQ
 YEGGVETLVFVRSYSTLNTYDVMVHPNGLIIVKFAHAGYISSAFEGTAQYQ
 NOYRENTLGTVTHSAHYDLDVGLKEMVMAEDMAFPTVPSAPDEQIOLQVTR
 KOLETEGQAAFLPGASPRYLIVLASKQKMGHPRGRYRIQTVSEFAPRPNQSTERA
 ISMGROIALVOTKRETPSSSVFNONDPTVPDADRIINNETTAGKRLVAVATGFA
 LHPHAEIDIPNTYVNGVGFELRPINFEDEPDSINSADSIYFQKHQDAGSECVNSLA
 CLEKDPACAPDLPAESHGGEFTN"

BASE COUNT 710 a 1070 c 1006 g 838 t
 ORIGIN

Query Match 47.9%; Score 1933.2; DB 4; Length 3624;
 Best Local Similarity 73.5%; Pident. No. 0; Mismatches 888; Indels 51; Gaps 9;
 Matches 2607; Conservative

```

Qy 11 CCCTTAGTCCAGCATCTCTACTACCGGAACCTCAGCCAGAGTCCGGAGCCGCCACC 70
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 45 CCCCCGGGCTGCAGAGATTGGCAGCAGAAAAACCTCAGCCAGAGTCCAGAAACCTCA-- 102
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 71 CCGTCCAGAGCAACAGAGCCCGCTTCTGCTGGCTGAGATACATTTGCTCTTTG 130
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 103 ---TACAGGGGCCAAAGAGAGACCTCTTCTGCTGGCTGAGAACACGCTTCCAGATTGGT 159
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 131 GTTAAATCAGTGTCCCTCTTCTGCGGAATAATGACAGAAAGACAAATCTGCTCTC 190
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 160 TCGATCTCTGTCTCTTAACACAGGACCTGTAACAGAAAGACACCTGCTGCTCTT 219
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 191 ATTCTGCGCGCATCACCATCTTGTGCTTGTGTCTGCTGCTGCTGAGGAGGGTGA 250
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 220 GCTTGCTGTGCATCACCACATCTTGTGCTGTGTGTGTCTTAATAGCTGAGGGAGA 279
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 251 GATGGGGGTGAACCAAGCAAGCTTCCCATTTGCTCTGTATTTCCAGTGCACACT 310
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 280 GATGGGGGTGAAGCAAGCAAGCTTCCCATTTGCTCTGCTGAGAACCCCAAGTGCAC 339
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 311 TGAACACACCTGGCCAGAGCCAGCTTTTGCACACCTGAGCCGAGAGAGAGCTAGCG 370
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 340 TGAACACACCTGGCCAGAGCCAGCTTTTGCACACCTGAGCCGAGAGAGAGCTAGCG 399
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 371 GTGATGCGCTTTTGACACCAAGCGGCTGGGGCAGAGGCTGTGATGACGCCAGCCCGG 430

```


Db	400	GTGATGACCTTCCTGACCCCAAGAACTGGGGCCAGACACTGGTGGATGACGCCACAGGCCCA	459
OY	431	CCCTGGCAACATGTGTCTTCCTCAGTGGAGTGTGCAGCTGCCTCCCAAGGCTGCAGCCCTG	490
Db	460	CCCTGTGACAACTCATCTCTTCTCGGTAGAGOTGACAGCTGGCCCCCAAGGCTGCAGCCCTG	519
OY	491	GCTACATTGGACAGAGGGGAGCCCCCACTGCCCCGGAGGACACTGGCCATGCTCTCTTT	550
Db	520	GCCCACTGGACAGAGGAGGCCCCCACTGCCCCGGAGGACACTGGCCATGCTCTCTTT	579
OY	551	GGCAGGCCAACCCAGCCCAACGTAGAGAGCTGGTGGGGGGCACAGCCTCACCCTTC	610
Db	580	GGTGGACAAACCCAGCCCAATGTACTGAGCTGGTGGGGGGCGCTGGCCCCAGCCCTTC	639
OY	611	TACATGGGGGACGTGACTGTGGAGCTGATAGAGAGCCCCCTGGCCTTACCCAGAGCCCC	670
Db	640	TACATGGGGGAGTGTGACCTGTGGAGACCTGATAGGGGGCCCCCTGGCCTTATCCAGAGCCCC	699
OY	671	GTCGCTTCCAAAGATACCTGAGCATAGACCATGATCTTCAACAGAGAGCTGGCCCCAG	730
Db	700	GTCGCTTCCCAAGAGTACCTGAGCATAGACCATGATCTTCAACAGAGAGCTGGCCCCAG	759
OY	731	GCTTCTGGGCTTCCACACACTGTGTGTTTACAAAGCAGCCGGGGAGCGAACCCTGGTGAC	790
Db	760	GCTGCTGGTGTCTGACACACTGTGTCTCTTCAACAAAGAGAGAGGAAACCTGGTGACC	819
OY	791	ATGACACAGGCTCCCCGCTGGTGTGCAATACAGGGGAGCCGGGCCAACCTGGTTGGACCTTAC	850
Db	820	ATGACACAGGCCCCCGGGGTTCGATACAGGTGACCCGGGCCAACCTGGTTGGACCTTAC	879
OY	851	TACAACTATCTCGGGCGCTGGGCTTCTCTGCAACACGCTGGGCTTGGAGCTGTAGTGAAAC	910
Db	880	TACAACTATCTAGGGCGCTGGGCTATCTCATCCCTGTGGGGTTAGAGCTGTGATAGAT	939
OY	911	CACAAGGCTTGTGACCTGTGCCGCTGGAGCTATCCAGAAAGGTGTTCTATCAAGCCGCTAC	970
Db	940	CACAAGGCTGTGAGACCTGTGCCAATGAGACATCCAGAAAGGTGTTCTTCAAGGCCGCTAC	999
OY	971	TACGACAGCCTGGGCCAGAGCTGAGAGGCCAGATTGTAGAGCCGGGCTGGATGTGGTGCTG	103
Db	1000	TATGAAAGTCTGGCCCAAGCTGAGAGAGCAATTGTAGAGCTGGCCGGGTGAATGTGGTGCTG	105
OY	1031	ATCCCAAGCAATGGACAGAGTGGGTCTGTCCCTGAAAGTCCCTGTGCCCCCGGGGTCCA	109
Db	1060	ATCCCAAAACATGGGACAGGAGGGGTCTGTGGCCCTGAAAGTCCCAAGGTGCCCGGGGTCCA	111
OY	1091	GCTGCCCTCTACAGTCTATCCCAAGGCCCGCGCTTCACTAGTGTCCAGGAGAGTCGAGTG	115
Db	1120	ACTCCACCTCTGCAAGTTCACATCCTCAGAGGAGACCGGCTTCACTAGTGTCCAGGAGACTCAGATG	117
OY	1151	GCTCCTCCTCAGTGGAGCTTCTCCTTGGGCTGGAGCATTCAGTGGGCCCAAGAGATCTTT	121
Db	1180	ACCTCCTCATTTGTGGACTTCTCTTGGGCTCGAGACTTTCAGTGGTCTCTAGAGATCTT	123
OY	1211	GACGTTGCTTCAAGAGGAAAGACTAGTTATGATTAAGCCTCCAAAGAGGCGTTGGCC	127
Db	1240	GACATTCGATTTCCAAAGGAAAGCACTGGCTTATGATACAGCCTGCAAGAGGCGGTGGCT	129
OY	1271	ATCATAGTGGAAATTTCCCAAGCAGCAATGACACCCGCTATGTGATGAGAGGCTTTGGC	133
Db	1300	ATTTATGTTGGGAATACCCCAAGCAGCAATGTCTCACTGTGCTATATGATGTGCTCTTTGGC	135
OY	1331	ATGGGCAAGTACACACAGGCCCTTACCCGGGGGTGAGAGCCCTTACCTTGGGACACTTAC	139
Db	1360	ATGGGCAAGTTCGCCACAGCCCTTACCCGAGGGGTGAGAGCTGCCCCCTTATCTGGGACACTTA	141
OY	1391	GTCGACTGGCACTTCTTTTGGATGCCAGGCCCCCAAGACAAATAGCTGATGCTTTTGT	145
Db	1420	GTCGACTGGCACTTCTTTTGGATGCTCAAGCCCCCAAGACCTACATGATGCTTTTGT	147
OY	1451	GTCGTTGAACGAACCAAGGCGTCCCCCTCGGGGACACCACTGATCTTACTGCGCAC	151

D	b	1480	GTGTTTAGCAGAAACAAGGGCCCTGCCCCCTGAGGGCCACACCACTCAGATTTTATTTTCCAG	1359
O	y	1511	TACTTTGGGGGCTCTTGCGGAAAGCGTGTGTCGTGATCATATCTATGTCACCTTGCTCAAC	1570
D	b	1540	TATTTTGGGGGCGTTTGAGAGACAGTGTGTGTCTTCAGATCTGTCTCAACTTGCTCAAC	1599
O	y	1571	TATGACTATGTGTGGATACAGTCTTCCACCCCAAGTGGGGCATAGAAATACATTTCTAT	1630
D	b	1600	TATGACTATGTGTGGATATGATGTCTTTCACCCCAATGGGGCATAGAAATCAATTCAT	1659
O	y	1631	GCCACGGGCTATACATCACTGGGATTCCTCTTGTGGTCACTGGGAATACGGGAACCA	1680
D	b	1660	GCCACGGGCTATATCACTATAGGGTCTCTTGTGGTCACTGGGAATACGGGAACCG	1719
O	y	1691	GTCACAGAGCACCCTGGGGCAGGTCCACACCAAGCGCCACTTCAAGGTGATGTC	1750
D	b	1720	GTTGGGAGAAATACACTGTGGGCAACCGTCCACACCAAGTGGCCACTTACAAAGTGAATG	1779
O	y	1751	GATGTACAGGACTGTGAGAACTGGTGTGGGCCGAGATATGTCTTGTGCCATGGCT	1810
D	b	1780	GATGTGGAGAGACTGTGAGAACTGGTGTGGGCCGAGAGCATGGCTTTGTGCCCAACG	1839
O	y	1811	GTCGCCGTGGAGCCCTGTAGCACACGTGCGAGAGCTGCAAGTCAACCCGGAACTGCTGAG	1870
D	b	1840	GTACCTCTGGAGCCCTGTAGCACACGATATACAGAGCTGTCAAGTCAACCCGGAACAGCTGAG	1899
O	y	1871	ATGAGAGAGCAGGGCCGCTCTCTGTGGGAAAGCGCACCCCTCGCTACTCTGATCGGAC	1930
D	b	1900	ACTGAGGAGCAGGGCGGCTTCCCTCGGAGAGGGGCTCCCTCGCTACTCTGATCGGAC	1959
O	y	1931	AGCAACCAAGCAACAAGTGGGGTCAACCCCGGGGCTACCGCATCAGATCTCAGCTT	1980
D	b	1960	AGCAAGCAGAGCAACAAGTGGGGGCAACCCCGGGGCTATCGCATCAGACAGCTCAGCTT	2019
O	y	1991	GCTGGAGAGCCGCTGCCCCCAAAACACTCTCATGTGCGAGAGGCTTCACTGGGAGAGTAC	2050
D	b	2020	GCTGGAGAGCCGCTGCCCCCAAAACACTCTCACAGAGAGCCATCAGTGGGGGAGGTAC	2079
O	y	2051	CAGCTGCTGTGACCCAGCGGAGAGAGAGAGAGCCAGTAGAGAGCGTTTCAATACAG	2110
D	b	2080	CAGCTGCTGTGACCCAGCCGAGAGAGAGAGAGAGAGCCAGTAGAGCGTTTCAATACAG	2139
O	y	2111	AATGACCTTTGGGGCCCACTGTGGAATTCAGTAGATTCATCAACATGAGACCAATTGCT	2170
D	b	2140	AATGACCCCTGAGACTCCCACTGTGGAATTCGTGATTCATCAACATGAGACCAATTGCT	2199
O	y	2171	GGAAGAAGTTTGGTGGGCTGGGTGACAGCTGTTTCTGTCAATCCCACTGTGAGAGGAC	2230
D	b	2200	GGAAGAAGCTTGGTGGGCTGGGTGACAGCGGTTTCTGTCAATCCCACTGTGAGAGGAC	2259
O	y	2231	ATTCCTTAACACAGTGTGTGGGGAACGGGCGTGGGCTTCTCTCGACCCCTATTAACATTC	2280
D	b	2260	ATTCCTTAACACAGTGTGTGGGGAACGGGTGTGGGCTTCTCTCGACCCCTACAACTTC	2319
O	y	2291	TTTGACGAAGACCCCTCTTCTACTGTGCGAGATCCATCTACTTCGAGGGGACCAAGAT	2350
D	b	2320	TTTGATGAGAGACCCCTCTATCAATTCGTGTGATCAATCTACTTCGAAACCAACAGAGAT	2379
O	y	2351	GCTGGGGCTGTGAGAGTCAACCCCTCAAGCTGTGCTGTGCCCAAGCTGTGCTGTGGCCCC	2410
D	b	2380	GCTGGGGCTGTGAGAGTCAACCTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2439
O	y	2411	GACTCTCTGCTCTCTCCACGGGGGCTTCTGTCAACATAGAGCGGTCTGTGAGTGGGAC	2470
D	b	2440	GATCTCTCTCTCTCTCCACGGGGGCTTCTGTCAACATTAAGGTCTTCTGTGGAATGGGA	2499
O	y	2471	ATGTGGCCAAGGGCTCCAGGGGCCAGGGGTGTGAGGGATGGGGAGCAGTGGGACTGGGCC	2530
D	b	2500	TATGTGACTGTGGGTCTAGGGGTCTGGGAGTGTGTGGGAGAGAGGCAAGAACTGGGGAACC	2559
O	y	2531	GGAGAGCTGGTCTCTTCTCTGTGTGCAAGAGACTCTTCTTCTCACTACCTCTCCCTGC	2580
D	b	2560	TGTGACCTTCTCTCTCTCTCAACACCCCTGTGGGCTCTCTCTCTCTCTCTCTCTCTCTCT	2619

QY 2591 ATCCGCTCTGACCCAGAGACCTCTGACCTGTATGCTGACACAGGGGACACTGAAAC 2650
 DB 2620 GCTTACCGGCTCACTGAGTATCATCGTATGCTGCTGATATGAGGGGCTTCTCAGC 2679
 QY 2651 CTGTGATGACGACGCTGATAGTATCATCATCAGAGAG-----CCAGGATGCGCCAGC 2706
 DB 2680 TCTGTGACACCCCAACTGCTGCTGCTTCTGCTGAGAGAGAGAGAGAGGCTCAGCCAAAGAG 2739
 QY 2707 CTGAGCCGCTGCGGAGGCTTC-----CCTAGAGT 2737
 DB 2740 CTGAGACTCAATGTTTACACGTTCTACAGACTCTCTGCTGTTGTTTTCATTTTCTT 2799
 QY 2738 GTTCCCTTGTGCTGCTGCTGCTTCCCAATTTTATAGCCACTCCCAAGACTCTA 2797
 DB 2800 TCTTTTGTCTTGTCTTCTGCTTCTCCCAATATGTTTATAGCCATGCTCCTCTG 2859
 QY 2798 AAGGGGCTATTCCCTGAGAGACCCAGAGAGTAGGTTGCCAGTCTGCAATGCTCATAGCTG 2857
 DB 2860 AGACTCCGACGCTCTCACTTGCACAGAGAGAGAGAGAGGCTCTCTGAGGATGATGCT 2919
 QY 2858 AGCTGAAAGATGCTTCTGCTCACAATTCCTCTCAATCCAGTCTTCTCTGCTT 2917
 DB 2920 AGAGCGCGGAGGAGGCTTCTGAGGCTTGAATCCAGAGAGTCCAGCTGGAAGGGG 2979
 QY 2918 CTTCTCTCTCACTACTTCT 2977
 DB 2980 TCTTTTGGCCATTTCT 3039
 QY 2978 TTTCTCCGCAATCTGAGGGGATATCCCTATGTCGCCAGCCCTGCTGACTCCCGAGCCT 3037
 DB 3040 CTCTCTCTCT--TCTTACCTATTTCTTATATACCTGGATTTACATCCAGCCCTGAGAGA 3097
 QY 3038 CAGTTTCACTCAAGTTCCGCTCTCTCTCTCCAGCCTATGAGAGTCTCAAGTCAAGGAG 3097
 DB 3098 TGTTCCTCAGCTCTCACTTTTCAACTGCTGCTCTCTCT--CCTGACTCTAGGCGCTGAGAA 3156
 QY 3098 CCTTATACAGAGGAGCAATCCCTGTGTGTGCTTCTCTGCTGTCTGTTATGAGAG 3157
 DB 3157 GTCTGAATATGCAAGGAGCATCTAGCTTGAAGAGAGAGTCCCTGTCTCTCTCTCTCT 3216
 QY 3158 TAGAGATGCT 3217
 DB 3217 TGGGAGCGGCGCTTCT 3275
 QY 3218 CATTTGTGTCT 3277
 DB 3276 CATTTGTATAGGCTCT--TCTTGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3333
 QY 3278 GCTCAAGTTCAATTTCAATGATTTGGGGTGGTCAATGATGATGATGATGATGATGAT 3337
 DB 3334 GCTCTATATGACACTTATAGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3393
 QY 3338 TACAGAGGCT 3392
 DB 3394 TGGCAGGT 3453
 QY 3393 GGGTGGCATCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3452
 DB 3454 GGGTGAAG 3511
 QY 3453 GGGTGGCATCAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3512
 DB 3512 GGGGAGCAATATGCTGTGAGCATGCAATGATGAGAGAGAGAGAGAGAGAGAGAGAG 3571
 QY 3513 GCGTAG 3518
 DB 3572 CCACAG 3577

RESULT 7
 AC046171/c
 LOCUS AC046171 189149 bp DNA 1linear HTG 03-MAR-2001

DEFINITION Homo sapiens chromosome 17 clone RP11-948G15 map 17, WORKING DRAFT
 AC046171
 VERSION AC046171.3 GI:13184209
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 189149)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 17, clone RP11-948G15
 Unpublished
 2 (bases 1 to 189149)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choe, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArnell, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferrer, P., Fitzhugh, N., Gage, D.,
 Galagan, J., Gardy, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehotzky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McNeeters, R.,
 Melidim, J., Menus, L., Mihova, T., Miranda, C., Miya, Y., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Plante, N.,
 Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange, R., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testa, S., Theodore, J., Tirrell, A., Travers, M., Tsigilis, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (13-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 2, 2001 this sequence version replaced gi:1712190.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7946
 Center clone name: 948.G.15
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 40% of reads
 Sequencing vector: Plasmid; n/a; 60% of reads
 Chemistry: Dye-terminator; Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 185448 bases at least Q40
 Consensus quality: 186995 bases at least Q30
 Consensus quality: 187536 bases at least Q20
 Insert size: 190000; agarose-IP
 Insert coverage: 187849; sum-of-coverage
 Quality coverage: 10.0 in Q20 bases; agarose-IP
 Quality coverage: 10.1 in Q20.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1116: contig of 1116 bp in length
 * 1117 11216: gap of 100 bp
 * 11217 13252: contig of 2036 bp in length

1. .21116

183243. .189149

42003 C 44237 g

```
servative 0; Mls
```

[illegible]

Dd	1612250	AGGATTTGGTGGCCTGGGTGACAGCTGGTATTTCTGCATATGCCACATGACAGAGCAATTC	1612191
Qy	2235	CTAACACAGTACAGTGTGGGGAACGGCGTGGGCTCTTCTCTCCGACCCTTAATACCTCTTGG	2294
Dd	1612130	CTAACACAGTACAGTGTGGGGAACGGCGTGGGCTCTTCTCTCCGACCCTTAATACCTCTTGG	1612131
Qy	2235	ACGAAGACCCCTCCTTCTACTCTGCGCACTCATCTAATTCGAGGGGACCAAGATGCTG	2254
Dd	1612130	ACGAAGACCCCTCCTTCTACTCTGCGCACTCATCTAATTCGAGGGGACCAAGATGCTG	1612071
Qy	2355	GGGGCTGGGAGGTAAACCCCTACTTGCCTGGCCCCAGGGGCGCTGCCTGGCCCCGAGC	2414
Dd	162070	GGGGCTGGGAGGTAAACCCCTACTTGCCTGGCCCCAGGGGCGCTGCCTGGCCCCGAGC	1620111
Qy	2415	TCCCTGCTTCTCCACAGGGGGCTTCTCTCAACACTAAGGGGTCTTGATGGAGGATGT	2474
Dd	162010	TCCCTGCTTCTCCACAGGGGGCTTCTCTCAACACTAAGGGGTCTTGATGGAGGATGT	1619511
Qy	2475	GGCCAAAGGGCTCCAGGGGCGAAGGGTGTAGGGATGGGAGACAGCTGGGGACTGGGGCGCA	2534
Dd	161950	GGCCAAAGGGCTCCAGGGGCGAAGGGTGTAGGGATGGGAGACAGCTGGGGACTGGGGCGCA	1618911
Qy	2535	GCCCTGTTCCCTCTTCTCTGTCGACAGACTCTCTTCTTCCACATCCCTCCCTGCATCC	2594
Dd	161890	GCCCTGTTCCCTCTTCTCTGTCGACAGACTCTCTTCTTCCACATCCCTCCCTGCATCC	1618311
Qy	2595	GCCCTTGAGCCAGAGACCTCCTGTACCTGTGTATGCTCTGACACAGGGGACACTGAACCTTG	2654
Dd	161830	GCCCTTGAGCCAGAGACCTCCTGTACCTGTGTATGCTCTGACACAGGGGACACTGAACCTTG	1617711
Qy	2655	TTGATGCCACGTGTACTAGTTCATCTACATCACAAGAGGGCAGAGCATGGGCCAGCCGAGGCC	2714
Dd	161770	TTGATGCCACGTGTACTAGTTCATCTACATCACAAGAGGGCAGAGCATGGGCCAGCCGAGGCC	1617111
Qy	2715	GTGGCCGAGGGCTTCCCTAGATGCTCCCTTGTGTGCTGTGTGGCTTTCGCCGAATCTTTT	2774
Dd	161710	GTGGCCGAGGGCTTCCCTAGATGCTCCCTTGTGTGCTGTGTGGCTTTCGCCGAATCTTTT	1616511
Qy	2775	TAGGCACTCTCCAAAGGACTCTAAAGGGGGGTATTCCTGGAGACCCCAAGATAGGGTTG	2834
Dd	161650	TAGGCACTCTCCAAAGGACTCTAAAGGGGGGTATTCCTGGAGACCCCAAGATAGGGTTG	1615911
Qy	2835	CCAGTCTCTGCAAGTCCATAGCTGAGCTGAGAAAGGATCTTCTGCTCAATCTCCCTCAT	2894
Dd	161590	CCAGTCTCTGCAAGTCCATAGCTGAGCTGAGAAAGGATCTTCTGCTCAATCTCCCTCAT	1615311
Qy	2895	CCAGGTCTCTTCTTCTTCTGCTTCTCTCTCTCTACACTTCTCTCTCTCTCTCTCTCTCTCT	2954
Dd	161530	CCAGGTCTCTTCTTCTTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1614711
Qy	2955	CTGCTTCTCTTCTATCTCGCAATTTCTCCGGAATCTGTAGGGGATATTCCTTAATGTCGCCA	3014
Dd	161470	CTGCTTCTCTTCTATCTCGCAATTTCTCCGGAATCTGTAGGGGATATTCCTTAATGTCGCCA	1614111
Qy	3015	GCCCCCTGCTACTCCCCCAGCCCTCTAGTTTTCAGTCAAGTTCGCTCTCTCTCCAGCCCTA	3074
Dd	161410	GCCCCCTGCTACTCCCCCAGCCCTCTAGTTTTCAGTCAAGTTCGCTCTCTCTCCAGCCCTA	1613511
Qy	3075	TGGAAGTCTCAAGGTCAAGGGGACCCCTAATCAGAGTGGCCAAATCCCTGTGTGTCCTGCC	3134
Dd	161350	TGGAAGTCTCAAGGTCAAGGGGACCCCTAATCAGAGTGGCCAAATCCCTGTGTGTCCTGCC	1612911
Qy	3135	TTGTGCTGTGCTTATTTGGGAGTAGAGATTGCTCTCAACCCCTGTGCTCTGGGGCTGGGTGT	3194
Dd	161290	TTGTGCTGTGCTTATTTGGGAGTAGAGATTGCTCTCAACCCCTGTGCTCTGGGGCTGGGTGT	1612311
Qy	3195	GTTTTCAGGACAGCTGCTTCTGTGATTTGTGTGCTGCTCTCATGTCTCTCTATAGAGGA	3254
Dd	161230	GTTTTCAGGACAGCTGCTTCTGTGATTTGTGTGCTGCTCTCATGTCTCTCTATAGAGGA	1611711
Qy	3255	GGATGTGTCATGTGACAGACAGCACTCAAGTTAGCAATTTCAAGTATTTGGGGGTGCAT	3314
Dd	161170	GGATGTGTCATGTGACAGACAGCACTCAAGTTAGCAATTTCAAGTATTTGGGGGTGCAT	1611111

QY 3315 GATATATGAAGATGGCCATTTTGTACAGAGGCTCTGTATTTCTGCAACAGCCTGTTTGGGA 3374
 |||||||
 Db 161110 GATATATGAAGATGGCCATTTTGTACAGAGGCTCTGTATTTCTGCAACAGCCTGTTTGGGA 161051
 |||||||
 QY 3375 GGCTGAGTGGAAACAAAGGTTGGGATCAAAAGATGAAAGCCAAAGCCCTTACACTCC 3434
 |||||||
 Db 161050 GGCTGAGTGGAAACAAAGGTTGGGATCAAAAGATGAAAGCCAAAGCCCTTACACTCC 160991
 |||||||
 QY 3435 AGCCACCCAGCAGAGAGGGCTGTCAATACATTCAGGATGCGAATGAGCTGGCCCT 3494
 |||||||
 Db 160990 AGCCACCCAGCAGAGAGGGCTGTCAATACATTCAGGATGCGAATGAGCTGGCCCT 160931
 |||||||
 QY 3495 GGGTGAAGTGGGGTCTGGCTTACTGGGAGAGGGCCCTGGCTGGGTGGGAGAGGCCCTGG 3554
 |||||||
 Db 160930 GGGTGAAGTGGGGTCTGGCTTACTGGGAGAGGGCCCTGGCTGGGTGGGAGAGGCCCTGG 160871
 |||||||
 QY 3555 CCTGTCCAGGCTTGGGCTCCATTCCTCATCTGCTCCTCGAGTGTGGATTGGG 3614
 |||||||
 Db 160870 CCTGTCCAGGCTTGGGCTCCATTCCTCATCTGCTCCTCGAGTGTGGATTGGG 160811
 |||||||
 QY 3615 GATGGGACAAAGAAATAGCAAGATGAGAACACAGAACTTTTCTCTAAAGGAC 3674
 |||||||
 Db 160810 GATGGGACAAAGAAATAGCAAGATGAGAACACAGAACTTTTCTCTAAAGGAC 160751
 |||||||
 QY 3675 TGGTAAATCAATCTGTATACAGCTTACATTAATAGTATGACGCTAAAAAATATATG 3734
 |||||||
 Db 160750 TGGTAAATCAATCTGTATACAGCTTACATTAATAGTATGACGCTAAAAAATATATG 160691
 |||||||
 QY 3735 TATGCTTATATCTATATATATATATCTTCAAGTGAAGGACAGCAGCAAGATG 3794
 |||||||
 Db 160690 TATGCTTATATCTATATATATATCTTCAAGTGAAGGACAGCAGCAAGATG 160631
 |||||||
 QY 3795 TGTATAGGCACTTCCCTTTGTTTTCAGAAAGAGTAGAATATAACACATAATTGCT 3854
 |||||||
 Db 160630 TGTATAGGCACTTCCCTTTGTTTTCAGAAAGAGTAGAATATAACACATAATTGCT 160571
 |||||||
 QY 3855 TATATAGCCATTCAGATTAATAGGCTTACATGATTTGGGAGGGAACCACTA 3914
 |||||||
 Db 160570 TATATAGCCATTCAGATTAATAGGCTTACATGATTTGGGAGGGAACCACTA 160511
 |||||||
 QY 3915 GGTGAGACAGAGAGAGGAGGCTTACACTTACACCTTTTGTACATTTGAATTT 3974
 |||||||
 Db 160510 GGTGAGACAGAGAGAGGAGGCTTACACTTACACCTTTTGTACATTTGAATTT 160451
 |||||||
 QY 3975 TGAACCATGTGCTATTTACCTATTCATAAATAAACAATAATGGCCCAAAAAA 4032
 |||||||
 Db 160450 TGAACCATGTGCTATTTACCTATTCATAAATAAACAATAATGGCCCAAAAAA 160393
 |||||||

RESULT 8
 AC055866/c 194567 bp DNA linear HTG 06-AUG-2002
 LOCUS Homo sapiens chromosome 17 clone RP11-376M2 map 17, *** SEQUENCING
 DEFINITION IN PROGRESS ***, 8 unordered pieces.
 ACCESSION AC055866
 VERSION AC055866.14 GI:22123743
 KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 194567)
 AUTHORS Birren,B., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 17, clone RP11-376M2
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 194567)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
 Campoliano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 COMMENT

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Laroque,K., Lamazeres,R., Landers,T., Lehouck,J.,
 Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
 McCarthy,M., McEwan,P., McGuck,K., McKernan,K., McPeeters,R.,
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testave,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 194567)
 Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choquel,Y., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gird,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nobu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Testave,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 6, 2002 this sequence version replaced gi:2131813.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 376_M_2
 Center clone name: 376_M_2
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 11855: contig of 11855 bp in length
 * 11856 11955: gap of 100 bp
 * 11956 15736: contig of 3781 bp in length
 * 15737 15836: gap of 100 bp
 * 15837 31925: contig of 16089 bp in length
 * 31926 32025: gap of 100 bp
 * 32026 148492: contig of 116467 bp in length
 * 148493 148592: gap of 100 bp
 * 148593 157642: contig of 9050 bp in length
 * 157643 157742: gap of 100 bp
 * 157743 167387: contig of 9645 bp in length
 * 167388 167487: gap of 100 bp

DEFINITION	Homo sapiens AOC3 gene for vascular adhesion protein-1, complete cds, exons 3 to 4.
ACCESSION	AB050502
VERSION	AB050502.1 GI:11602472
KEYWORDS	
SEGMENT	
SOURCE	3 of 3 Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:keio BAC library clone:KB551F. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
ORGANISM	
REFERENCE	1 Inamura,Y., Kudo,H., Masihira,Y., Oguchi,Y. and Shimizu,N. Tandem array of two human copper amine oxidase genes (AOC2 and AOC3) in 17q21 Unpublished 2 (bases 1 to 3233) Shimizu,N.
JOURNAL	Direct Submission Submitted (23-OCT-2000) Nobuyoshi Shimizu, Keio University School of Medicine, Department of Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@db.med.keio.ac.jp, Tel:03-3351-2370, Fax:03-3351-2370)
AUTHORS	
TITLE	
JOURNAL	
FEATURES	Location/Qualifiers 1..3233 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="17" /map="17q21" /clone="KB551F" /cell_line="FLEB14-14" /cell_type="pre-Pro-B cell" /clone_id="Keio BAC library" join(AB050500.1:639..2238, AB050501.1:581..866,556..685, 1387..1662) /gene="AOC3"
SOURCE	join(AB050500.1:639..2238, AB050501.1:581..866,556..685, 1387..1662) /gene="AOC3" /EC_number="1.4.3.6" /note="homolog of semicarbazide-sensitive amine oxidase placental monamine oxidase" /codon_start=1 /product="vascular adhesion protein-1" /protein_id="BAB18866.1" /db_xref="GI:11602473" translation="MNOKTLLIIVITFAVCYLVLVGRGGDGGSPSLPHCPSPS SPSKQPTHTPGSQSLPADLSREETAVNRRTORLGRPELVDAARAPRDNCVSVELCL LPFPAAADLAHLDRGSPPARERALELTVFFRQRPQPVSLVGLPRLPHSTMDYVERKIG GGPIYHRPRVLFQEYLDIDQMIFNRELTPQASGLIHGCCFKHGSRMLVTYTAPRG OSGRATVFGLYVINISGAGFLAHVGELALVNHRALPAPMTIKVFQGYYSLAQG LEAFGEAGLVNVILIPDNGTGSMSLKSPVPGPAPPLQFYPOGPRFSVOGSRAVSI WTFSEFGAGFSGPRIIPDYRFQGERLYEISLOEALAIYGNSPAMTTRVYDGEGMGMO KYTPLRNGVDCCPLATYVDVNHFLLESAPRTITDARCFYPQONCIPRRHSDLYSX YFGEAEVLVVRSMSTLANLDYWDVTFHFSGAIEIRFATGITSSFLTGATKGIG NQVEHNTLGTVHTSHAKKVDLDVAAGLENVMWAEDMFWPAVWPSPHQLRQLQVTH KLIMEEQAAFLVGSAAEPFRYLTLASNHNKMGHGRGYIQWLISFAGEPLPNSSMARQ ESMRVYLATVQRKEEPSSESSSVNONDPAPATVDSDFINNETIAGDELVAWTAGAG LIHTHAEDIPTVTVGNCGFELRPRIYNFDDBPSYSDSYIFRGDDQAACEVNPPLPL CLPDAACAPDLPAPFSSHGFESHN"
gene	<1..553 /gene="AOC3" 556..685 /gene="AOC3" /number=3 686..1386 /gene="AOC3" 1387..3233 /number=4
CDS	BASE COUNT 694 a 844 C 854 g 841 t
ORIGIN	

		Best Local Similarity 100.0%; Pred. No. 0;	
		Matches 1845; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	2175	AGGATTTGGTGGCCCTGGGTGACAGCTGGTTTTTCTGCATATCCACATGCAGAGACATTC	2234
Db	1385	AGGATTTGGTGGCCCTGGGTGACAGCTGGTTTTTCTGCATATCCACATGCAGAGACATTC	1444
QY	2235	CTAACACAGTACTGTGGGGGAACGGCGTGGGGCTTCTTCTCTCGAACCCCTAATCTTGTG	2294
Db	1445	CTAACACAGTACTGTGGGGGAACGGCGTGGGGCTTCTTCTCTCGAACCCCTAATCTTGTG	1504
QY	2295	ACGAAGACCCCTCCCTTCTACTGTGCGAGTCCATCTACTTCCGAGGGGACCGAGATGCTG	2354
Db	1505	ACGAAGACCCCTCCCTTCTACTGTGCGAGTCCATCTACTTCCGAGGGGACCGAGATGCTG	1564
QY	2355	GGGGCTGGAGGTCAACCCCTAGCTTCCTGTCGCCACAGCTGCTGCTGTGCCCCGAC	2414
Db	1565	GGGGCTGGAGGTCAACCCCTAGCTTCCTGTCGCCACAGCTGCTGCTGTGCCCCGAC	1624
QY	2415	TCCCTGCTTCTCCACGCGGGCTTCTCTCACACTAGCGGCTCTGGGATGGGCACTG	2474
Db	1625	TCCCTGCTTCTCCACGCGGGCTTCTCTCACACTAGCGGCTCTGGGATGGGCACTG	1684
QY	2475	GGGCAAGGGCTCCAGGGGCGAGGGGTGAGGGGTGGGGAGACATGGGCACTGGGCGGGA	2534
Db	1685	GGGCAAGGGCTCCAGGGGCGAGGGGTGAGGGGTGGGGAGACACTGGGCACTGGGCGGGA	1744
QY	2535	GCCCTGTTCCCTCTTCTTCCCTGTCACAGACTCTCTTCTTCCACTACCCCTCCGCATCC	2594
Db	1745	GCCCTGTTCCCTCTTCTTCCCTGTCACAGACTCTCTTCTTCCACTACCCCTCCGCATCC	1804
QY	2595	GCCCTGAGCCAGAGACCTCTGACCCCTGTGATGCTTGACACAGGGGACACTGAACCTTG	2654
Db	1805	GCCCTGAGCCAGAGACCTCTGACCCCTGTGATGCTTGACACAGGGGACACTGAACCTTG	1864
QY	2655	TTGATGCCAGCTGTACTAGTTCGATCCACAGAGGACGAGTGGGCCGACGCTGGAGGC	2714
Db	1865	TTGATGCCAGCTGTACTAGTTCGATCCACAGAGGACGAGTGGGCCGACGCTGGAGGC	1924
QY	2715	GTGGCGGAGGGCTTCCCTAGATGTTCCCTTGTGTGCTGTGCGTTTCCCAATCTTTT	2774
Db	1925	GTGGCGGAGGGCTTCCCTAGATGTTCCCTTGTGTGCTGTGCGTTTCCCAATCTTTT	1984
QY	2775	TAGGCCACCTCCAAAGACTCTAAAAGGGGGCTATTCCCTGGAGACCCAGATAGGGTTG	2834
Db	1985	TAGGCCACCTCCAAAGACTCTAAAAGGGGGCTATTCCCTGGAGACCCAGATAGGGTTG	2044
QY	2835	CCAGTCTCTGCAGTCCATAGCTGAGCTGGAAAGAGTCTTGTCTCACATTCCTCTCAT	2894
Db	2045	CCAGTCTCTGCAGTCCATAGCTGAGCTGGAAAGAGTCTTGTCTCACATTCCTCTCAT	2104
QY	2895	CCAGGTCCCTTCTCTCTCTGCTTCTCTCTCTCTCACCTACTTCTCTCTCTCTCTCTGTC	2954
Db	2105	CCAGGTCCCTTCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGTC	2164
QY	2955	CTGCTTCTCTTCTATCTGCATATTTCTCCGAATCTGAGGGGATATCCTATGTCCA	3014
Db	2165	CTGCTTCTCTTCTATCTGCATATTTCTCCGAATCTGAGGGGATATCCTATGTCCA	2224
QY	3015	GCCCTGTGACTCCCGCAGCCCTCAGTTTTCAGTCAAGTTCGGTCTCTCTCTCCAGCCCA	3074
Db	2225	GCCCTGTGACTCCCGCAGCCCTCAGTTTTCAGTCAAGTTCGGTCTCTCTCTCCAGCCCA	2284
QY	3075	TGGAAGTCTCAAGGTACAGGGGACCCCTATTCAGAGTGGGCAATCCCTGTGTCGTTCC	3134
Db	2285	TGGAAGTCTCAAGGTACAGGGGACCCCTATTCAGAGTGGGCAATCCCTGTGTCGTTCC	2344
QY	3135	TGTGTCTGTGCTATTGGGATAGAGATGCTCTTACCCTGTCTCTGGGCTGGGTGT	3194
Db	2345	TGTGTCTGTGCTATTGGGATAGAGATGCTCTTACCCTGTCTCTGGGCTGGGTGT	2404
QY	3195	GTTTAGAGACAGCTGCTTCTGTGCAATTTGTCTGCTGCTCTCATGCTCTATAGAGA	3254

Db	2405	GTTCAGGACAGCTGCTCTGTGCAATTTGTGTGCTGCGCTCAGCTCTCTATAGAGA	2464
QY	3255	GGATGCTCATCGTGCACAGCAGCAGCTCAAGTTAGCATTTCAAGTATTTGGGGGTGCAT	3314
Db	2465	GGATGCTCATCGTGCACAGCAGCAGCTCAAGTTAGCATTTCAAGTATTTGGGGGTGCAT	2524
QY	3315	GATTAATGAAGAATGGCCATTTTGTATACAGGGGCTCTGTATTTCTGCAACAGCCTGTTGGGA	3374
Db	2525	GATTAATGAAGAATGGCCATTTTGTATACAGGGGCTCTGTATTTCTGCAACAGCCTGTTGGGA	2584
QY	3375	GGCTGAGTGTGAAACCAAAAGGCTGGGCATCAAAAGATGAAAGGCCAAAGCCCTCACACTCC	3434
Db	2585	GGCTGAGTGTGAAACCAAAAGGCTGGGCATCAAAAGATGAAAGGCCAAAGCCCTCACACTCC	2644
QY	3435	AGCCACCCAGCCAGAGAGGGGCTGTCCAACTCACTCAGGCATGCGAATGAGCTGGGCCCT	3494
Db	2645	AGCCACCCAGCCAGAGAGGGGCTGTCCAACTCACTCAGGCATGCGAATGAGCTGGGCCCT	2704
QY	3495	GGGTGAGGTGGGGGTCTGCGGCTAGGGGGAGGGGGCTGGCCGGTGGGGGCAAGGGCCCTGG	3554
Db	2705	GGGTGAGGTGGGGGTCTGCGGCTAGGGGGAGGGGGCTGGCCGGTGGGGGCAAGGGCCCTGG	2764
QY	3555	CCTGGTCCAGGCTTGGGCTCCATTCCTCACTCACTGCTGCTCCCTCCTGAGGTCGGAATTGGG	3614
Db	2765	CCTGGTCCAGGCTTGGGCTCCATTCCTCACTCACTGCTGCTCCCTCCTGAGGTCGGAATTGGG	2824
QY	3615	GATGGGGCAAAAGAAATAGCAAGAGATAGAAACAACAGAAATTTTCTCTTAAAGAC	3674
Db	2825	GATGGGGCAAAAGAAATAGCAAGAGATAGAAACAACAGAAATTTTCTCTTAAAGAC	2884
QY	3675	TGGTTAAATCAATTCTGTACAGCCTTCAATPACAATGTATGCGCTAAAAAATAATTG	3734
Db	2885	TGGTTAAATCAATTCTGTACAGCCTTCAATPACAATGTATGCGCTAAAAAATAATTG	2944
QY	3735	TATGCTTTATATACATAATATGTATAATCTTCAGGTGAAAAAGGCCACAGAAATG	3794
Db	2945	TATGCTTTATATACATAATATGTATAATCTTCAGGTGAAAAAGGCCACAGAAATG	3004
QY	3795	TGTAATAGGCACTTCCCATTTGTGTGTAGAAAGAGTAGAATATAAACATAATTGCT	3854
Db	3005	TGTAATAGGCACTTCCCATTTGTGTGTAGAAAGAGTAGAATATAAACATAATTGCT	3064
QY	3855	TATGTATGCTCAATTCAGAAATAATGGGTAAACACTGATTACTTTGGAGGGGACACAGTA	3914
Db	3065	TATGTATGCTCAATTCAGAAATAATGGGTAAACACTGATTACTTTGGAGGGGACACAGTA	3124
QY	3915	GGTTGAGGACAGAGAGAGGAAGGCTTTAAACACTTACACCCCTTTGTACATTTGAAATT	3974
Db	3125	GGTTGAGGACAGAGAGAGGAAGGCTTTAAACACTTACACCCCTTTGTACATTTGAAATT	3184
QY	3975	TGAACCAATGACGTATTTACCTATTCAAAATTAACATAAATGGGGCC	4023
Db	3185	TGAACCAATGACGTATTTACCTATTCAAAATTAACATAAATGGGGCC	3233

RESULT	10
Bty15774	
LOCUS	Bty15774
DEFINITION	Bos taurus mRNA for copper amine oxidase.
ACCESSION	J15774
VERSION	J15774.1
KEYWORDS	G1:2665357
SOURCE	copper amine oxidase. cow.
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
AUTHORS	Bovidae; Bovinae; Bos. 1 (bases 1 to 2605) Hogdall,J.,E.V., Houen,G., Borre,M., Bundgaard,J.R., Larsson,L.I. and Vusture and tissue-specific expression of genes encoding bovine copper amine oxidases
TITLE	
JOURNAL	Eur. J. Biochem. 251 (1-2), 320-328 (1998)

Medline	98151264
PubMed	9492300
Reference	2 (bases 1 to 2605)
Authors	Vuust, J.
Title	Direct Submission
Journal	Submitted (03-DEC-1997) J. Vuust, Statens Serum Institut, Artiklerives 5, DS-2300 Copenhagen, DENMARK
Comment	Related sequence: 127218.
Features	Location/Qualifiers
Source	1. 2605
gene	/organism="Bos taurus"
	/db_xref="taxon:9913"
	/sex="Female"
	/tissue_type="lung"
	/clone_id="Clontech"
	/dev_stage="adult"
	12..2300
	/gene="BOLA0"
	12..2300
	/EC_number="1.4.3.6"
	/codon_start=1
	/product="copper amine oxidase"
	/protein_id="CAA5776.1"
	/db_xref="GI:2663358"
	/db_xref="SPRMBL:O46406"
	/translation="MFIFILSLMTLVMGREGVSGEEGVKCOHPSLPKPCPSRS PSDPPTHPDOSOLFDSREBELTAVMSGLTQKPLDYDAARPSDKITFVEIDL PPKAAHLADRPPARRALATVFFGGQFQVTELVGLPFPSPYMDVVERGG GLPLRYRHTVLREYLDIDQMFIRELPQAAGVLDHCSSYKGGGMLVMTTPAPGQ SGRATFPGLYYNISKAGVYLHPVGLLVDHKAIDPAQMTIQVFQGRVYSVLQ EEOEGRAVNVVLPNNGTSGMSLSKQVPPGPPPLQFPGDGRFSGOBSRTSLM TFSGLGAFSGSPRIEFDIQRQERLAYRISLOEVAIYGMTPAAMLTIRYDAQGMCK FAPLTGRCVDCPYLATYVDMLHLESAPRTLDARFVFQNGKPLRRHHSPIISQ FGVYEVYLVRSYSTILNLDYMDVYFHNGALVKAHVTGTISAFEGGTAKQYEN QVRENTLGTVTHRSAAHKVDLDVGGLENMWAEDMAVPTVWSPSPHOIQVRIK OLEEEDQAFPLPAGGSGSRPYLYLASKGNKMGHPRGYIQVTSFAGRLPONSTERRI SMGWYOLATVQRKTEBSSSSSVFENONDPMTFADFIINNETIACKDLVAMTYAPLAC HIPAEIDPMTVYGVNGVGFELRPYNFDDPDSMDSISYFREGDQASCEINPLAC LPOAQTCAPLPVPVSHGCPYR"
BASE COUNT	543 a 760 c 734 g 568 t
ORIGIN	
Query Match	44.2%; Score 1787.6; DB 4; Length 2605;
Best Local Similarity	82.5%; Pred. No. 0;
Matches 2065; Conservative	0; Mismatches 429; Indels 10; Gaps 1;
QY	221 GTTGTGTCCTGCTGCTGGTGGGAGGGGTGGAGATGGGGGTGAACCCAGCCAGCAGCTTCCCAT 280
DB	69 GGTGTGTTGGGAGAGAGAGAGAGAGATTGGGAAGCAATGTCAATCCAGCCGCTCCCGC 128
QY	281 TGCCTCTGTATATCTCCCAAGTGCCAGGCTTGAGACACACCTGGCCAGAGCCAGCTGTT 340
DB	129 TGCCTCTCAGATCCCTCAGTAGACAGCCCTGGACACACCTGGACCAAGCCACTCTGTT 188
QY	341 GCAGACTGAGCCGAGAGAGAGAGCTGAGCGCTGTGATGACCTTTTGACCCAGCCAGCGTGGG 400
DB	189 GCAGACTGAGCCGAGAGAGAGAGCTGAGCTGTGATGACCTTCTTGACCCAGAGCTGGGG 248
QY	401 CCAGGCTGCTGTGATGACAGCCAGGCCGCCGCTTCGACCAACTGTCTTCTCAAGTGGAG 460
DB	249 CCAGACTGCTGTGATGACAGCCAGGCCGCCGCTTCGACCAACTGTCTTCTCGGTAGAG 308
QY	461 TTGCAGCTGCTCCCAAGGCTGACAGCCCTGGCTCACTTGGACAGAGGGAGAGCCCCACCT 520
DB	309 CTGCAGCTGCTCCCAAGGCTGACAGCCCTGGCTCACTTGGACAGAGAGAGAGCCCCACCT 368
QY	521 GCCCGGAGGAGCACTGGCCATGCTCTTCTTTGGACAGGCAACCCCAAGCCAGCTGAGTAG 580
DB	369 GCCCGGAGGAGCACTGGCCATGCTCTTCTTTGGTGGACACACCCCAAGCTGAGTAG 428
QY	581 CTGTGTGTGGGAGCACTGCTTACCCCTCTTACATGGCGGAGCTGATGTTGGAGCGTCAAT 640

Db 429 CTGTGTGTGGGGCCGCTGCCCAAGCCCTCTACATGCGGGATGTACGCTGAGCTCAT 488
Qy 641 GGAGGCCCCCTGCTCCCTATCACCAGAGCCCGCTGCTGTCCAGAGTACCTGACATAGAC 700
Db 489 GGGGGCCCCCTGCTCCCTATCACCAGAGCCCGCTGCTGTCCAGAGTACCTGACATAGAC 548
Qy 701 CAGATGATTTTCAACAGAGAGCTGCCCGCTCTGCGGCTCTCCACCATTCTTCTTC 760
Db 549 CAGATGATTTTCAACAGAGAGCTGCCCGCTGCTGTCCAGAGTACCTGACATAGAC 608
Qy 761 TCAAGACACCGGGAGAGAGCTGTGATACATACCCAGGCTCCCGTGTGCAATCA 820
Db 609 TCAAGACACCGGGAGAGAGCTGTGATACATACCCAGGCTCCCGTGTGCAATCA 668
Qy 821 GGGGACCGGGCCACTGTGTTGGCTCTACTACATCTCGGGCGTGGGTTCTTCTG 880
Db 669 GGTGACCGGGCCACTGTGTTGGCTCTACTACATCTCAAGGCTGGGTTCTTCTG 728
Qy 881 CACGACGTGGGCTTGGAGTGTGATGACCAAGGCGCTTGAACCTGCGCGCTGGACT 940
Db 729 CACGCTGTGGGTTAGAGCTGTGTGATGACCAAGGCTTGGAGCTGCGCGCTGGACT 788
Qy 941 ATCCAGAGGTGTTCTATCAAGGCGCTCTACTACAGAGCTGCGCGCTGAGGCGCAG 1000
Db 789 ATCCAGAGGTGTTCTTCAAGGCGCTCTACTATGAAGTGTGCTCAGCTGAGAGCAG 848
Qy 1001 TTTGAGCGCGCGCTGTGATGTGTGTGTATCCAGACATGTGGACAGTGGTCTGTG 1060
Db 849 TTTGAGCGCGCGGTGTGATGTGTGTGTATCCAGACATGTGGACAGTGGTCTGTG 908
Qy 1061 TCCCTAAGTCCCTGTGCGCGCGGGTCCAGCTCCCGCTTACAGTTCTTCCCAAGC 1120
Db 909 TCCCTAAGTCCCAAGTGTGCGCGCGGGTCCAGCTCCAGCTGTGAGTTCTTCAAGGC 968
Qy 1121 CCCCCTGAGTGTCCAGAGGAGTGTGAGTGTGCTCTCACTGTGAGTTCTTCTTGGC 1180
Db 969 ACCCGTGTGAGTGTCCAGAGGAGTGTGAGTGTGCTCTCACTGTGAGTTCTTCTTGGC 1028
Qy 1181 CTGGGAGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1240
Db 1029 CTGGGAGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1088
Qy 1241 TATGATATTAAGCTTCCAGAGGCGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1300
Db 1089 TATGATATTAAGCTTCCAGAGGCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1148
Qy 1301 ACAGCCGCTATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1360
Db 1149 CTGACTGTGATATGATGTGCTGTGCTTGGCATGGGCAAGTTGGCGACGCCCTGACCGA 1208
Qy 1361 GGGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1420
Db 1209 GGGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1268
Qy 1421 GCGCCCAAGACATAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1480
Db 1269 GCGCCCAAGACCTACTATGTGCTTGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1328
Qy 1481 GCGCGACACCATGATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1540
Db 1329 AGCGACACCATGATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1388
Qy 1541 GTGTGATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1600
Db 1389 GTGTGATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1448
Qy 1601 CCCAGTGGGCGATAGAAATACGATTTATGAGGCGCTATACCTGAGCTGGGCTTCTG 1660
Db 1449 CCCAGTGGGCGATAGAAATACGATTTATGAGGCGCTATACCTGAGCTGGGCTTCTG 1508
Qy 1661 TTTGTGTCTACTGGGAAGTACGGGAGACCAAGTGTGAGAGCACACCTGGGCGAGGCTCAC 1720
Db 1509 TTTGTGTCTACTGGGAAATAGGAAACCAAGTGTGGGAGAAATACACTGGGCGAGGCTCAC 1568

Qy 1721 ACCCAGAGCGCCCACTTCAAGGTGATCTGTGATGTGAGAGAGTGTGAGTGTGAGTGTG 1780
Db 1569 ACCCAGAGTGTGCGCACTTCAAGGTGATCTGTGATGTGAGAGAGTGTGAGTGTGAGTGTG 1628
Qy 1781 GCGGAGATATGATGTTTGTGCGCATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1840
Db 1629 GCTGAGAGATATGATGTTTGTGCGCATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1688
Qy 1841 AGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1900
Db 1689 AGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1748
Qy 1901 AGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1960
Db 1749 GGGGTTTCCCTGTGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1808
Qy 1961 CCGGCTTACCGCATCCAGATGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2020
Db 1809 CCGGCTTACCGCATCCAGATGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1868
Qy 2021 ATGGCAGAGGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2080
Db 1869 ACAGAGAGGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1928
Qy 2081 GAGCCCAATGAGCAGAGCGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2140
Db 1929 GAGCCCAATGAGCAGAGCGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1988
Qy 2141 AGTACTTGTATCAACATGAGAGCATTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2200
Db 1989 GCTGACTTGTATCAACATGAGAGCATTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2048
Qy 2201 GGTGTTGTGATATCCCATGAGAGCATTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2260
Db 2049 GGTGTTGTGATATCCCATGAGAGCATTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2108
Qy 2261 GTGGGCTTGTGCTGCGAGCCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2320
Db 2109 GTGGGCTTGTGCTGCGAGCCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2168
Qy 2321 GACTCCATCTACTTCCGAGGAGGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2380
Db 2169 GACTCCATCTACTTCCGAGGAGGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2228
Qy 2381 TGCTGTGCGGAGCTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2440
Db 2229 TGCTGTGCGGAGCTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2288
Qy 2441 TGTCAACAATGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2500
Db 2289 CTTGAGTACTAGTGTGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2348
Qy 2501 GAGGATGTGAGGAGCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2560
Db 2349 GTGAGAGGAGGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2398
Qy 2561 GACTCTCTTCTTCTGACTTACCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2620
Db 2399 GTTCTCTCTCTGACTTACCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2458
Qy 2621 CTGTGATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2680
Db 2459 CTGTGATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2518
Qy 2681 TCCACAGAGGCGAGGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2724
Db 2519 TCCGAGAGAGAGAGAGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2562

RESULT 11
AF054985
LOCUS AF054985 1794 bp mRNA linear PRI 02-APR-1998

DEFINITION Homo sapiens clone 23577 placenta copper monamine oxidase mRNA, partial cds.

ACCESSION AF054985.1 GI:3005693

VERSION AF054985.1

KEYWORDS

SOURCE Homo sapiens.

ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1794)
Anderson, B., Wentland, M.A., Ricafrente, J.Y., Liu, W. and Gibbs, R.A. A 'double adaptor' method for improved shotgun library construction 96207227

JOURNAL Anal. Biochem. 236 (1), 107-113 (1996)

PMID 8619474

REFERENCE 2 (bases 1 to 1794)
Yu, W., Anderson, B., Morley, K.C., Muzny, D.M., Ding, Y., Liu, W., Ricafrente, J.Y., Wentland, M.A., Lennon, G. and Gibbs, R.A. Large-scale concatenation cDNA sequencing

JOURNAL Genome Res. 7 (4), 353-358 (1997)

PMID 9110174

TITLE 97264341

JOURNAL MEDLINE 9110174

PUBMED 3 (bases 1 to 1794)

REFERENCE Yu, W. and Gibbs, R.A. Direct Submission

AUTHORS Submitted (23-MAR-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA

JOURNAL

FEATURES

source

1..1794
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE Consortium 23577"
/sex="female"
/tissue_type="brain"
/clone_lib="INIB"
/dev_stage="infant"
1..196
/note="similar to Homo sapiens placenta copper monamine oxidase encoded by Genbank Accession Number U39447"

CDS

BASE COUNT 397 a 1482 c 455 g 460 t

ORIGIN

Query Match 44.0%; Score 1779.2; DB 9; Length 1794;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1781; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2257 CGGCGTGGGCTTCTTCCTCCGACCTATTAATCTTTGACGAAGACCCCTCTTACTC 2316
1 CGGCGTGGGCTTCTTCCTCCGACCTATTAATCTTTGACGAAGACCCCTCTTACTC 60

2317 TGCGACATCAATCTACTTCCGAGGGGACAGATGCTGGGGCTGGAGGTAAACCCCT 2376
1 TGCGACATCAATCTACTTCCGAGGGGACAGATGCTGGGGCTGGAGGTAAACCCCT 120

61 TGCGACATCAATCTACTTCCGAGGGGACAGATGCTGGGGCTGGAGGTAAACCCCT 120

2377 AGCTTGCCCTGCCAGGCTGCTGCTGTCGCCAGCACCCTGCTTCCACAGGGGG 2436
1 AGCTTGCCCTGCCAGGCTGCTGCTGTCGCCAGCACCCTGCTTCCACAGGGGG 180

121 AGCTTGCCCTGCCAGGCTGCTGCTGTCGCCAGCACCCTGCTTCCACAGGGGG 180

2437 CTTCCTCAACAATAGGCGGCTCTGGATGGGATGTGGCCAAAGGGCTCCAGGGCCAG 2496
1 CTTCCTCAACAATAGGCGGCTCTGGATGGGATGTGGCCAAAGGGCTCCAGGGCCAG 240

181 CTTCCTCAACAATAGGCGGCTCTGGATGGGATGTGGCCAAAGGGCTCCAGGGCCAG 240

2497 GTGTAGGAGATGGGAGAGAGCTGGGGCTGGGGCCGAGCCTGCTTCTTCTGTG 2556
1 GTGTAGGAGATGGGAGAGAGCTGGGGCTGGGGCCGAGCCTGCTTCTTCTGTG 300

241 GTGTAGGAGATGGGAGAGAGCTGGGGCTGGGGCCGAGCCTGCTTCTTCTGTG 300

2557 CCAGACTCTCTTCTTCCACTACCTCCCTCGATCGGCTGTGAGCAGAGGCTCT 2616

301 CCAGAGCTCTTCTTCTTCCACTACCTCCCTCGATCGGCTGTAGCCAGGAGCTCTCT 360
2617 GACCTGTGATGCTGACACAGAGGAGACATGAACCTTGTGATGACAGTACTAGAT 2676
361 GACCTGTGATGCTGACACAGAGGAGACATGAACCTTGTGATGACAGTACTAGAT 420

2677 CTATTCACAGAGGAGCAGGATGCCCCAGCCTGAGACCGCTGGCCAGGCTTCCATAG 2736
421 CTATTCACAGAGGAGCAGGATGCCCCAGCCTGAGACCGCTGGCCAGGCTTCCATAG 480

2737 GGTTCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2796
481 GGTTCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

2797 AAAGGGGCTATTCCTGAGAGACCCAGAGATGAGGTTGGCACTCCGCAAGTCAAGT 2856
541 AAAGGGGCTATTCCTGAGAGACCCAGAGATGAGGTTGGCACTCCGCAAGTCAAGT 600

2857 GAGCTGAAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2916
601 GAGCTGAAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660

2917 TCTCTCTCTACACTTCT 2976
661 TCTCTCTCTACACTTCT 720

2977 ATTTCCTCCGATCTCTGAGGAGATATCCCTATGTCACAGCCCTGTAATCCAGGCC 3036
721 ATTTCCTCCGATCTCTGAGGAGATATCCCTATGTCACAGCCCTGTAATCCAGGCC 780

3037 TCAGTTTCAGTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3096
781 TCAGTTTCAGTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

3097 CCCCAATCAGAGTGGGCAATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3156
841 CCCCAATCAGAGTGGGCAATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900

3157 GTAGAGATGCTCTACACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3216
901 GTAGAGATGCTCTACACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960

3217 GCATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3276
961 GCATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020

3277 ACCTCAAGTTAGCATTTCAAGATTTGGGGGTGCAATGATATGAAAGATGGCATTTT 3336
1021 ACCTCAAGTTAGCATTTCAAGATTTGGGGGTGCAATGATATGAAAGATGGCATTTT 1080

3337 GTACGAGGCTCTGATTTCTGCAAGAGCTGTTTGGAGGCTGAGGTGAAACAAAGGT 3396
1081 GTACGAGGCTCTGATTTCTGCAAGAGCTGTTTGGAGGCTGAGGTGAAACAAAGGT 1140

3397 GGGCATCAAAAGATGAGAAAGCAAGCCCTTACAACTCCAGCCACCCAGCAGAGGGCT 3456
1141 GGGCATCAAAAGATGAGAAAGCAAGCCCTTACAACTCCAGCCACCCAGCAGAGGGCT 1200

3457 GTCCATCATCATTCAGAGCATGGAATGAGCTGGGGCTGGGGGTGAGTGGGGGTGGCT 3516
1201 GTCCATCATCATTCAGAGCATGGAATGAGCTGGGGCTGGGGGTGAGTGGGGGTGGCT 1260

3517 AGTGGGAGGGGCTGAGCTGAGGTGGGGAGGGGCTGAGCTGAGCTGAGCTGAGCTGCA 3576
1261 AGTGGGAGGGGCTGAGCTGAGGTGGGGAGGGGCTGAGCTGAGCTGAGCTGAGCTGCA 1320

3577 TTCCCATCATCT 3636
1321 TTCCCATCATCT 1380

3637 GAGATGAGAAACAGAAACTTTTCTCTAAAGAGCTGGTTAATGTAATCTCATACA 3696

Db 1381 GAGATGAGAACACAGAACTTTTCTGTTAAAGGACTGTTAAATCAATTCATACA 1440

QY 3697 GCCTTACATACATAGTATGACAGCTAAATAATATGTATGCTTTATATCTAATATG 3756

Db 1441 GCCTTACATACATAGTATGACAGCTAAATAATATGTATGCTTTATATCTAATATG 1500

QY 3757 TAAATATCTTACAGTGAAGAAAGGCAAGCCACAGAAATGTATAGCGACTTCCATTGG 3816

Db 1501 TAAATATCTTACAGTGAAGAAAGGCAAGCCACAGAAATGTATAGCGACTTCCATTGG 1560

QY 3817 TGTTCACAAGGAGTAAATATTAACATATGCTTATATGCTTATGAGCAATTA 3876

Db 1561 TGTTCACAAGGAGTAAATATTAACATATGCTTATATGCTTATGAGCAATTA 1620

QY 3877 ATGGCTAACACTGATTAATTTTGGAGGGAACCAAGTATGAGAGACAGAGAGGAAG 3936

Db 1621 ATGGCTAACACTGATTAATTTTGGAGGGAACCAAGTATGAGAGACAGAGAGGAAG 1680

QY 3937 GGTCTTAACACTTACACCTTTTGTACATTTTGAATTTTGAACCATGTGACTGTATACC 3996

Db 1681 GGTCTTAACACTTACACCTTTTGTACATTTTGAATTTTGAACCATGTGACTGTATACC 1740

QY 3997 TATTCATAATATACATTAATATGAGCCCAACAGAAAAA 4040

Db 1741 TATTCATAATATACATTAATATGAGCCCAACAGAAAAA 1784

RESULT 12

AB05050051 2614 bp DNA linear PRI 05-DEC-2000

DEFINITION Homo sapiens AOC3 gene for vascular adhesion protein-1, exon 1.

ACCESSION AB050500

VERSION AB050500.1 GI:11602470

KEYWORDS

SEGMENT

SOURCE 1 of 3

ORGANISM Homo sapiens pre-pro-B cell cell_line: FLEB14-14 DNA, clone_11b: Keio

REFERENCE

AUTHORS Immamura, Y., Kudoh, J., Mashima, Y., Oguchi, Y. and Shimizu, N.

TITLE Tanden array of two human copper amine oxidase genes (AOC2 and AOC3) in 17q21

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2614)

AUTHORS Shimizu, N.

TITLE Direct Submission

JOURNAL Submitted (25-OCT-2000) Nobuyoshi Shimizu, Keio University School of Medicine, Department of Molecular Biology, 35 Shinanomachi, Shijuku-ku, Tokyo 160, Japan (E-mail: shimizu@med.keio.ac.jp, Tel: 03-3351-2370, Fax: 03-3351-2370)

FEATURES

source

1. 2614

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="17"

/map="17q21"

/clone="KB51F9"

/cell_line="FLEB14-14"

/cell_type="pre-pro-B cell"

/clone_11b="Keio BAC library"

/note="SP-1 site"

432..437

/note="SP-1 site"

453..458

/note="SP-1 site"

479..2238

/gene="AOC3"

479..2238

/gene="AOC3"

/product="vascular adhesion protein-1"

/note="CDS is reported in Acc#: AB050502.

homolog of semicarbazide-sensitive amine oxidase

placental monoamine oxidase"

/number=1

/EC_number="1.4.3.6"

2239..2614

/gene="AOC3"

BASE COUNT 516 a 757 c 732 g 609 t

ORIGIN

Query Match 43.6%; Score 1761; DB 9; Length 2614;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTCCACACCTTATGTCACAGCATCATACCGGAACTCAGCCAGATCCGGGA 60

Db 479 GTCTTCCACACCTTATGTCACAGCATCATACCGGAACTCAGCCAGATCCGGGA 538

QY 61 GCCCCCAACCCCGTCCAGAGGCAACAGAGAGCCCGTCTTGTGCGGTGAGATCATTTG 120

Db 539 GCCCCCAACCCCGTCCAGAGGCAACAGAGAGCCCGTCTTGTGCGGTGAGATCATTTG 598

QY 121 CTCTCTTGTGTTAATCAGCTGTCCCTCTTGTGGAATAATGAACAGACAAATCCT 180

Db 599 CTCTCTTGTGTTAATCAGCTGTCCCTCTTGTGGAATAATGAACAGACAAATCCT 658

QY 181 CGTCTCTCATTTTGGCCGTCATCAGCATCTTGGCTTGTTTGTGTCTGCTGGTGG 240

Db 659 CGTCTCTCATTTTGGCCGTCATCAGCATCTTGGCTTGTTTGTGTCTGCTGGTGG 718

QY 241 CAGGGGTGAGATGGGGGTGAACCCAGCAGCTTCCCATTCGCTGTATCTCCAG 300

Db 719 CAGGGGTGAGATGGGGGTGAACCCAGCAGCTTCCCATTCGCTGTATCTCCAG 778

QY 301 TGCACAGCCTTGAGACACACCTGCGCAGACCAAGCTGTTTGCAGACTGAGCCGAGAGA 360

Db 779 TGCACAGCCTTGAGACACACCTGCGCAGACCAAGCTGTTTGCAGACTGAGCCGAGAGA 838

QY 361 GCTGACGCTGTGATGCGCTTCTGACCCAGAGGCGTGGGGCCAGGCGTGGTATCAGC 420

Db 839 GCTGACGCTGTGATGCGCTTCTGACCCAGAGGCGTGGGGCCAGGCGTGGTATCAGC 898

QY 421 CCAGGCGCGGCGCTCGAGACATGTCTTCTGAGTGAAGTTGACAGCTGCTCCCAAGC 480

Db 899 CCAGGCGCGGCGCTCGAGACATGTCTTCTGAGTGAAGTTGACAGCTGCTCCCAAGC 958

QY 481 TGCAGCCTTGCTCACTTGAGACAGGGGAGACCCCACTGCCCGGAGGCACTGGCCAT 540

Db 959 TGCAGCCTTGCTCACTTGAGACAGGGGAGACCCCACTGCCCGGAGGCACTGGCCAT 1018

QY 541 CGTCTCTTTTGGAGGCAACCCAGCCCAAGCTGAGTGAAGTGGTGGGGCACTGGC 600

Db 1019 CGTCTCTTTTGGAGGCAACCCAGCCCAAGCTGAGTGAAGTGGTGGGGCACTGGC 1078

QY 601 TCACCCCTCTATACGCGGAGCTGAGTGAAGCTGATGAGGAGCCCTTGCCTATTA 660

Db 1079 TCACCCCTCTATACGCGGAGCTGAGTGAAGCTGATGAGGAGCCCTTGCCTATTA 1138

QY 661 CCAGAGCCCGGCTGCTGTCCAGAGTACTGAGCATAGACAGATGATCTTCAACAGAGA 720

Db 1139 CCAGAGCCCGGCTGCTGTCCAGAGTACTGAGCATAGACAGATGATCTTCAACAGAGA 1198

QY 721 GCTGCCCCAGGCTTGTGGCTTCTCACACACTGTGCTTTACAGAGCAGGGAGGGA 780

Db 1199 GCTGCCCCAGGCTTGTGGCTTCTCACACACTGTGCTTTACAGAGCAGGGAGGGA 1258

QY 781 CTTGTGACATGACACAGGCTCCCGTGTGTCATGACATGAGGAGCCGACCACTGTT 840

Db 1259 CTTGTGACATGACACAGGCTCCCGTGTGTCATGACATGAGGAGCCGACCACTGTT 1318

QY 841 TGGCTCTTACTACAACTTCGGGCGCTGGGTTCTTCTCCACACAGCTGGGCTTGAAGCT 900

Db 1319 TGGCTCTTACTACAACTTCGGGCGCTGGGTTCTTCTCCACACAGCTGGGCTTGAAGCT 1378

QY 901 GCTAGTGAACCAAGGCGCTTGAACCTGCCCGCTGACTATATCCAGAAAGTGTCTATCA 960

|||||
Db 1379 GCTAGTACACAGAGCCCTTGACCTGCGCCCTGACTATCCAGAGGCTGCTATCA 1438
QY 961 AGCGCGCTACTACAGACAGCCCTGCGCCAGCTGAGAGCCAGTTTGAAGCCGCGCTGTGAA 1020
Db 1439 AGCGCGCTACTACAGACAGCCCTGCGCCAGCTGAGAGCCAGTTTGAAGCCGCGCTGTGAA 1438
QY 1021 TGTGCTGCTATCCAGACATAGGACAGTGGGTCTGCTGCTGAACTCCCTGTGCC 1080
Db 1499 TGTGCTGCTATCCAGACATAGGACAGTGGGTCTGCTGCTGAACTCCCTGTGCC 1558
QY 1081 CCCGGGTACAGCTCCCTCTACAGTTCTATCCCAAGGCCCCCGCTCACTGCTCCAGG 1140
Db 1559 CCCGGGTACAGCTCCCTCTACAGTTCTATCCCAAGGCCCCCGCTCACTGCTCCAGG 1618
QY 1141 AAGTCAGTGGCCCTCTACCTGTGAGACTTCTCTCTTGGCCCTGAGACATCACTGAGCC 1200
Db 1619 AAGTCAGTGGCCCTCTACCTGTGAGACTTCTCTCTTGGCCCTGAGACATCACTGAGCC 1678
QY 1201 AAGATCTTTGACCTTGTGCTTCCAGAGAAAGACTGTTATAGATAGCTCCAGGA 1260
Db 1679 AAGATCTTTGACCTTGTGCTTCCAGAGAAAGACTGTTATAGATAGCTCCAGGA 1738
QY 1261 GACCTGACCATATGATGTAATCCCGACAGCAATGACACCGCTATGATGG 1320
Db 1739 GACCTGACCATATGATGTAATCCCGACAGCAATGACACCGCTATGATGG 1798
QY 1321 AGCTTTGGCATGGGCAAGTACACACGCCCCCTGACCGTGGGTGATGCTCCCTACTT 1380
Db 1799 AGCTTTGGCATGGGCAAGTACACACGCCCCCTGACCGTGGGTGATGCTCCCTACTT 1858
QY 1381 GGGCAGCTAGTGGACAGCTCTCTTTGAGATCCAGGCCCCCAAGACATAGCTGA 1440
Db 1859 GGGCAGCTAGTGGACAGCTCTCTTTGAGATCCAGGCCCCCAAGACATAGCTGA 1918
QY 1441 TGCCTTTGCTGTTTGAACAGAACAGGCTCCCTGCGGAGACACACTAGATCT 1500
Db 1919 TGCCTTTGCTGTTTGAACAGAACAGGCTCCCTGCGGAGACACACTAGATCT 1978
QY 1501 CTACTGCACTACTTTGGGGCTCTTGGGAAACGGTGTGCTCTAGATCTATGCTCAC 1560
Db 1979 CTACTGCACTACTTTGGGGCTCTTGGGAAACGGTGTGCTCTAGATCTATGCTCAC 2038
QY 1561 CTTCCTCACTACTATGATGTTGGGATAGGGCTTCCACCCCAAGTGGGGCATAGAAAT 1620
Db 2039 CTTCCTCACTACTATGATGTTGGGATAGGGCTTCCACCCCAAGTGGGGCATAGAAAT 2098
QY 1621 ACATCTTATGCCAGCGGCTACATCAGCTGCGCATCTCTTTGGTCTACTGGAAGTA 1680
Db 2099 ACATCTTATGCCAGCGGCTACATCAGCTGCGCATCTCTTTGGTCTACTGGAAGTA 2158
QY 1681 CGGAAACCAAGTGTACAGACACACCTGGGACGGTCCACACCCACAGCCCACTTCAA 1740
Db 2159 CGGAAACCAAGTGTACAGACACACCTGGGACGGTCCACACCCACAGCCCACTTCAA 2218
QY 1741 GGTGATCTGGATGTAGCAGG 1761
Db 2219 GGTGATCTGGATGTAGCAGG 2239

RESULT 13
BOVRA 2664 bp mRNA linear MAM 03-AUG-1994
LOCUS Bos taurus serum amine oxidase mRNA, complete cds.
DEFINITION L27218.1 GI:520784
ACCESSION L27218.1
VERSION 1
KEYWORDS amine oxidase.
SOURCE Bos taurus female liver, bovine cDNA to mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 2664)

AUTHORS Nu,D., Medzhiradzky,K.F., Adams,G.W., Mayer,P., Hines,W.M.,
TITLE Burlingame,A.L., Smith,A.J., Cai,D., and Kliman,J.P.
JOURNAL Primary Structures for Mammalian Intracellular and Serum Copper
FEATURES J. Biol. Chem. (1994) In press
SOURCE Location/Qualifiers
1. 2664
/organism="Bos taurus"
/db_xref="taxon:913"
/sex="female"
/tissue_type="liver, bovine"
61..2349
CDS
61..2349
/codon_start=1
/product="serum amine oxidase"
/protein_id="AA30525.1"
/db_xref="GI:520785"
/translation="MFIFLSLWLLVNGREGVSGEGVKOCHPSLPSPRS
PSPQPTTHPQSLFADLSRELLTTHSTLQDGLDPAQARSDKVSVELQ
PKRAALAHLDGSPPPAREALATVEFGQPOPNVETLVVGPJPSPYRVDYVERHG
GPLPYRRPVLRLREYLDIDOMFNRLPQAGVLAHHCSSYKQGGKLTMSNAPRGVQ
SGDRSTWFGIYVITKGGPYLHPVGLLELDVHKALDPAQVQKVFQGRYENLAOL
EEOFEAGOVNVVVIIPDDGTGFWSLKSOVPGPTPLQHPQGRPREPVQNGRYASLM
TFSEGLGAFSGPRFDFVREGELAYEISLOEAGAYAGNTPAAMLTMYDSFGMGY
FAPLIRGVDCPYLATYMDMHPVSESQTKLHDACVPEQNKGLPLRRHSFSLSHY
FGVAQTVLFRSVSTMLNDYMDVETFNGLIEVLAHATGITSALFLGARIRGN
QVEHTLGPVHTSHAHKYVDLDVGLNWMVAEDMAFPTALPWSPEHIOILQVTRK
OLETEBOAEFLPGASPRYLYLASKOSNMGHEDGRIOTVSPAGPMQNSMERAF
SMGRYOLAIQREKETESPSSSVYNQNDPPTVDFEFDLINETIACKDLVAWTTAGFL
HIPHARDIPMTVTVGVGVGFELPYNFPQEPESMDSADSIYFREGDASCEINPLAC
LPQATCADLPVFSHGYPKY"
BASE COUNT 549 a 781 c 751 g 583 t
ORIGIN
Query Match 43.4%; Score 1754; DB 4; Length 2664;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 450; Indels 10; Gaps 1;
QY 221 GTTGTGCTCTGCTGCTGGGAGGGGTGAGAGTGGGGTGAACCCAGCTTCCCAT 280
Db 118 GTTGTGCTGAGTGAAGGAGGAGTGGGAAGCAATGTATCCACCTGCTCCCGC 177
QY 281 TGCCCTCTGATCTCCAGTGGCCAGCTTGGACACACCTGGCCAGAGCAAGCTTT 340
Db 178 TGCCCTCTCAGATCTCCATGTAGACACACCTTGACACACCTTGACACACCTTT 237
QY 341 GCAGACCTGAGCCGAGAGAGCTGACGCTGTGATGCGCTTTTGACCCAGCGCTGGG 400
Db 238 GCAGACCTGAGCCGAGAGAGCTGACAACTGTGATGACTTCTGACTAGCAGCTGGG 297
QY 401 CCAGAGCTGTGTGATGAGAGCCAGGCCCCGCTCGGACAACTGTCTCTCAGTGAG 460
Db 298 CCAGAGCTGTGTGATGAGAGCCAGGCCCCGCTCGGACAACTGTCTCTCGGTAGAA 357
QY 461 TTGAGCTGCTCCCAAGGCTGACAGCCCTGAGCTAGTGAACAGGGGAGAGCCGCCACT 520
Db 358 CTTCAGCTGCCCCCAAGGCTGACAGCCCTGAGCCACTGACAGAGGAGAGAGCCGCCACT 417
QY 521 GCCCGGAGAGCATGGCCATGCTCTTGTGCAAGGCAACCCAGCCCAAGTGAAGTGA 580
Db 418 GCCCGGAGAGCATGGCCATGCTCTTGTGCGGACCAACCCAGCCCAAGTGAAGTGA 477
QY 581 CTGTGTGTTGGGGCCACATGCTCTACACCTCTCTACATGGGGAGACGTGATGGAGCTCAT 640
Db 478 CTGTGTGTTGGGGCCGCTGCCCCCAGCCCTCTACATGGGGAGTGAAGCTGGAGCGCAT 537
QY 641 GGAAGCCCTGCGCTATCAACGAGCCCGCTGCTGTTCAGAGATACCTGACATAGAC 700
Db 538 GGGCGCCCCCTGCGCTATTCACAGAGCCCGCTGCTTCGAGAGATACCTGACATAGAC 597
QY 701 CAGATGATCTTCAACAGAGAGTGGCCCAAGGCTTGTGGGCTTCTCCACCACTGTGCTTC 760
Db 598 CAGATGATCTTCAACAGAGAGTGGCCCAAGGCTGCTGTGCTCTCCACCACTGTGCTTC 657


```

REFERENCE      1 (bases 1 to 2664)
AUTHORS        Mu, D., Medzhradszky, K.F., Adams, G.W., Mayer, P., Hines, W.M.,
                Bullingame, A.L., Smith, A.J., Cai, D. and Klimmen, J.P.
TITLE          Primary structures for a mammalian cellular and serum copper amine
                oxidase
JOURNAL        J. Biol. Chem. 269 (13), 9926-9932 (1994)
MEDLINE        94193686
PUBMED        8144587
REMARK         GenBank staff at the National Library of Medicine created this
                entry [NCBI gidsq 146050] from the original journal article.
                This sequence comes from Fig. 3.
FEATURES
  source        1..2664
                location/Qualifiers
  gene          /organism="Bos taurus"
                /db_xref="taxon:9913"
                1..2664
  CDS           /gene="copper amine oxidase, BSAO"
                61..2349
                /note="copper amine oxidase, BSAO"
                /note="amiloride-binding protein homolog; precursor;
                Method: conceptual translation with partial peptide
                sequencing; This sequence comes from Fig. 3; BSAO"
                /codon_start=1
                /product="copper amine oxidase"
                /protein_id="AAB30397.1"
                /db_xref="GI:546216"
                /translation="MRFIFLSLMTLLVMREBQCVSEBVGKQCHPSLPKPSRSS
                PSDPHTHPDSQLEFADLSREELTVMSFTLQDLPDLVAAQAPDNCVSEYELQ
                PKRAALAHLDKGGPPARALALVFEGSGPQPAVTELVVGLPQPMRYVBERG
                GLPLRYRPLVLELYLDIDIMFNRELPOQAQVHHCSSKQGGKLLTMMASRGVQ
                SGRSTFGIYVNTKGGPYLLPYGLLELVDKALDPADMTYOKVFPQGRYEMLAOL
                EOEPAQVWVWVVPDGGTGGFMSKSOVSPGPPPLPHQGRSPFVQGRVASSLM
                TFSRGLAFSGPVPVDFQGRERLAEISLOBAGAVYGNTPAMLTRHMSDLSHY
                FAPPLIKGVDCPLIATYMDHMFVEYSQTPRLHDAFCVFEDNKLPLRRHSDLSHY
                FGAAGTVALFVSSTMLNDYWDVYFNGALAEVLAHATGYISFLKGAARYN
                OVEHTGLPVHTSAHKYVDLVGLENWMAEDMAVPAIPIPSPHOQLORLVTK
                QLETEEGAPFLGASPRYLLASKQNMKGHPGYRIQTVSFQGGPMQPSPERAF
                SMGRYOLATPQRETEPSSSSVSENONDPMTPTGVFSPFINNETAGDLYAMVAGFL
                HIPHAEDIPNTVWNGVGFELRPNFEDDEPMSDLSIYFRGQDAGSEINPLAC
                LPOATTCADPLPVSNGHYPEY"
BASE COUNT     549 a      781 c      751 g      583 t
ORIGIN
Query Match    43.4%: Score 1754; DB 4; Length 2664;
Best Local Similarity 81.6%: Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 450; Indels 10; Gaps 1;

```

```

Db 478 CTGATGATGGGGCCGCTGCCCCAGCCCTCTACATGCGGAGTGTGACCGGTGACGTCTAT 537
Qy 641 GGAGGCCCCCTGCTCCCTATACCGAGAGCCCGGTGCTTCCAAAGATACCTGACATAGAC 700
Db 538 GGGGGCCCCCTGCTCCCTATACCGAGAGCCCGGTGCTTCCGAGATACCTGACATAGAC 597
Qy 701 CAGATGATCTTCAACAGAGAGCTGCCCGAGCTTCTGGGCTTCTCCACCACTGTTGCTTC 760
Db 598 CAGATGATCTTCAACAGAGAGCTGCCCGAGCTTCTGGGCTTCTCCACCACTGTTGCTTC 657
Qy 761 TACAAGCAGCGGGGAGGAGACCTGTGTGACATGACACAGGCTCCCGGTGCTGATATCA 820
Db 658 TACAAGCAGGAGGAGACAAACTGTGTGACATGAACTGATCCCGGTGAGATGACATCA 717
Qy 821 GGGGACCGGGCCACCTGTTGGCTTGTACTACAAACATCTCGGGGCTGGGCTTCTCTG 880
Db 718 GGTATATGTTCACTGTTGGCTTGTACTACATATACATACAAAGGTTGGGCTTCTCTG 777
Qy 881 CACCACTGTTGGCTTGTACTGATGACCAACAAAGCCCTTGACCTGCCCCGCTGGACT 940
Db 778 CACCCCTGTTGGCTTGTACTGATGACCAATAGGCTTGGACCCCTGGACCTGGATTGGAC 837
Qy 941 ATCCAGAGGTGTTCTATCAAGCCCGCTACTACAGCAGCCGCGCCACCTGGAGCCCGAC 1000
Db 838 GTCCAGAGGTGTTCTTTCAGAGCCCGCTACTATGAAATCTGGCCACCTGGAGAGCAG 897
Qy 1001 TTGAGGCGCGGCTGCTGAATGTGGTCTGATCCAGAAATGGCAGAGTGGTCTG 1060
Db 898 TTGAGGCGCGGCTGCTGAATGTGGTCTGATCCAGAAATGGCAGAGTGGTCTG 957
Qy 1061 TCCCTGAAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1120
Db 958 TCCCTGAAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017
Qy 1121 CCCCCTCACTGTCACAGAGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 1180
Db 1018 CCCCCTCACTGTCACAGAGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 1077
Qy 1181 CTCGAGCATTCAGTGGCCCAAGGATCTTTAGCGTTCCTTCCAGAGAAAGACTAGTT 1240
Db 1078 CTCGAGCATTCAGTGGCCCAAGGATCTTTAGCGTTCCTTCCAGAGAAAGACTAGTT 1137
Qy 1241 TATGATATAGCTTCACAGAGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1300
Db 1138 TATGATATAGCTTCACAGAGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1197
Qy 1301 ACGAGCCCTATGTGAGAGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 1360
Db 1198 CTCACCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1257
Qy 1361 GGGTGGAGTCCCTTACTTGGCCACCTAGCTGGAAGTGGACCTTCTTGGAGTCCAG 1420
Db 1258 GGGTGGAGTCCCTTACTTGGCCACCTAGCTGGAAGTGGACCTTCTTGGAGTCCAG 1317
Qy 1421 GCCCCAAGACATACGATGATGCTTTGGTGTGTAAGCAAGAGGGCTCCCTG 1480
Db 1318 ACCCCAGACATACATGATGATGCTTTGGTGTGTAAGCAAGAGGGCTCCCTG 1377
Qy 1481 CGGAGACACATACATGATGCTTTGGTGTGTAAGCAAGAGGGCTCCCTG 1540
Db 1378 AGGGAGACACATGATGATGCTTTGGTGTGTAAGCAAGAGGGCTCCCTG 1437
Qy 1541 GTGCTAGATATATGCTACCTGCTCAATATGATATGATGATGATGATGATGATGAT 1600
Db 1438 GTGCTAGATATATGCTACCTGCTCAATATGATATGATGATGATGATGATGATGAT 1497
Qy 1601 CCCAGTGGGGGACATAGAAATACGATTTCTATGACCAAGGGCTTCAATACCTGCGCAT 1660
Db 1498 CCCAGTGGGGGACATAGAAATACGATTTCTATGACCAAGGGCTTCAATACCTGCGCAT 1557
Qy 1661 TTTGCTGCTACTGAGAGTGGAGCAACAAAGTGCAGAGCACCTTGGAGGACGCTGCAC 1720
Db 1558 TTTGCTGCTGCTGAGAGTGGAGCAACAAAGTGCAGAGCACCTTGGAGGACGCTGCAC 1617

```

QY	1721	ACCACAGGCCCCACATTCAAGGATGATCTGATGTAGACAGACTGGGAACCTGGGCTGTG	1780
Db	1618	ACCCACAGTGTCCCACTACAAAGGTGATCTGTAGACTGGGAGGACTGGAAAACTGGGTCTGG	1677
QY	1781	GCCGAGGATATGGTCTTTGTGCCATGTGCTGTGCCCTGTGAGCCCTGTGAGCCAGCTTCAG	1840
Db	1678	GCTGAGGATATGTGCTTTTGTGCCAGGCGATACCCCTGTGAGCCCTGTGAGCCAGCTTCAG	1737
QY	1901	AGCGCCACCCCTGTGATACCTGTACCTGTGGCCAGCAACCAACAGCAAGTGGGTCACCCC	1960
Db	1798	GGGGCTTTCCCTGTGCTTACCTGTGTCTGTGGCCAGCAGAGCAAGCAAGTGGGTCACCCCT	1857
QY	1961	CGGGGCTTACCGCATTCAGATGCTCAGCTTTGCTGTGAGAGCGCGCTGCCCAAAACAGCTCC	2020
Db	1858	CGCGGCTATGTGATTCAGACCCGTTCAGCTTTGTGGGGGGCGGATGCCCAAGAACGCCCC	1917
QY	2021	ATGGCGAGAGGCTTTCAGCTGGGAGAGGTACCAGCTGGCTGTGACCCAGCGGAAGAGAGG	2080
Db	1918	ATGAGAGAGACCTTTCAGCTGGGAGAGGTACAGCTGGCATTAACCCAGAGAGAGAGACA	1977
QY	2081	GAGCCGATGTAGCAGCAGCTTTTCAATCAGAAATGACCTTTGGGCCCCACATGTGATTTTC	2140
Db	1978	GAGCCGATGTAGCAGCAGCTTTTCAATCAGAAATGACCTTTGGGCCCCACATGTGATTTTC	2037
QY	2141	AGTGACTTTCATCAACATGAGACCATTTCTGTGAAAAGATTTGTGTGCTGGGTGACAGCT	2200
Db	2038	TCTGACTTTCATCAACATGAGACCATTTCTGTGAAAAGATTTGTGTGCTGGGTGACAGCC	2097
QY	2201	GGTTTCTGCATATCCCAATGAGAGACCACTTCTTAACACAGATCATGTGGGGAAGAGGC	2280
Db	2098	GGTTTCTGCATATCCCAATGAGAGACCACTTCTTAACACAGATCATGTGGGGAAGAGGC	2157
QY	2261	GTGGGCTTTCCTCCGACACCTATATACCTTTTACGAGAGACCCCTCTTCTACTCTGCC	2330
Db	2158	GTGGGCTTTCCTCCGACACCTATATACCTTTTACGAGAGACCCCTCTTCTACTCTGCC	2217
QY	2321	GACTTCATCTACTTCCGAGGGGACAGAGATGCTGGGGCTGTGAGATCAACCCCTACT	2380
Db	2218	GACTTCATCTACTTCCGAGGGGACAGAGATGCTGGGGCTGTGAGATCAACCCCTACT	2277
QY	2381	TGCGTGGCCAGGCTGTGCGCTGTGCCCCGACCTCCCTGTGCTTCTCCACAGGGGGCTTC	2440
Db	2278	TGCGTGGCCAGGCTGTGCGCTGTGCCCCGACCTCCCTGTGCTTCTCCACAGGGGGCTTC	2337
QY	2441	TCTCACAACATGTGGGCTCTGGGATGTGGGCAATGTGGCCAAAGGGCTCCAGGCCAGAGGT	2500
Db	2338	CCTGAGTACTAGGTGTGCTTGTGGGGCAGAGATGTGTAGAGACCCAAAGAGAGAGT	2397
QY	2501	GAGGAGTGGGAGACAGCTGGGCACTGGGCGGAGACCTGTGTTCCCTCTTCTGTGGCAG	2560
Db	2398	GTTGAGAGAGGGGAGGGGCTGGGCACTGTGAG-----TTTCTCTCCACACTCCACCCAG	2447
QY	2561	GACTCTCTTTTTCACATACCTCCCTGTGACCTGTGACCCCTGTGAGCCAGAGAGCTCTGTAGC	2620
Db	2448	GTTCTCTCTCTCTCCCATTTTGTGCCCTTGTCTCTCAACCCCTGTGAGAGATTCATATAGC	2507
QY	2621	CTGTGATGCTGTGACACAGGGGACATGTGACCTTGTGTGATGTGCGACAGCTGTAGTTCACA	2680
Db	2508	CTGTGATGCTGTGACACAGGGGCTTGTCAACTGTGTGACTCCTGTGACCTGTGATTCCTTA	2567
QY	2681	TCCACAGAGCCAGGCACTGGCCCAAGCTGTGAGAGCGGTGGCCGAGG	2724
Db	2568	TCCGAGAGAGAAAGATGTGCCAAGCTGGGAGTCTAGATGTG	2611

DEFINITION	Mus musculus semicarbazide-sensitive amine oxidase mRNA, complete cds.
ACCESSION	AF115411
VERSION	AF115411.1
KEYWORDS	GI:4185816
SOURCE	
ORGANISM	Mus musculus.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognatha; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 4210)
TITLE	Moldes,M., Fève,B. and Pairault,J.
JOURNAL	Molecular cloning of a major mRNA species in murine 3T3 adipocyte lineage: differentiation-dependent expression, regulation, and identification as semicarbazide-sensitive amine oxidase
MEDLINE	J. Biol. Chem. 274 (14), 9515-9523 (1999)
PUBMED	99194802
REFERENCES	2 (bases 1 to 4210)
AUTHORS	Moldes,M., Fève,B. and Pairault,J.
TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-1999) Institut Biomedical des Cordeliers, UPRESA 7079 UPMC CNRS, 15,rue de l'Ecole de Medecine, 75270 Paris Cedex 06, France
FEATURES	
source	location/Qualifiers
	1..4210
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/cell_line="3T3-L1"
	166..2463
	/EC number="1.4.3.6"
	/function="catalyzes the oxydation of primary amines such as benzylamine, methyl amine, N1-acetyl-putrescine, beta phenylethylamine, tyramine and histamine
	/note="transmembrane enzyme; strongly inhibited by semicarbazide, aminoguanidine and hydrazine derivatives such as phenelzine and hydrazine; abundantly expressed in white and brown adipose tissues and the aorta; expressed at a low level in lung and muscle; MSOAO"
	/codon_start=1
	/product="semicarbazide-sensitive amine oxidase"
	/protein_id="AAD09199.1"
	/db_xref="GI:4185817"
	/translation="MCKTFLVLLALAVITAFALVCVLVLAGSGDGGLSPLPCPSVLPSPQPTTHSSQSPRADISPEELTAYMSPLTKGLGIVDAQAPSDNVEVELLGLPAAALAHLDHGRGPPVRBALAIIFFGGQPKRNVSELYVGPLPHPSYRDVYERHGGPPLRYRRLVDREYQDIEMIFHRLPQASGLHHCCKFHGGONLTLTAPRLQSGGRATAPFGLYVMSGAGFPHPHIGELIDHKALPALMTIDKRVYQGSYSLTQLEDFEAGLVNVVLPNNGTGSWSLKSVPGPAPLPQFHPQCPKRSVQSGVSSSLMAFSPFGAFSGSPRIPIIDIRQGERVAEISVQEAIALYCGNSPASMGCYDVGFGIGKSPILRGKDCPLATYVDMHFLLEQAPRTLLDACVPEQONGLPRRHSHSPFSGHYFGGVGTVLVRVSVSTLNTLDYIMDWPFHNGAIEYKFNATGIISSAFPGAEKKGNRGAHILGTVHTSHAFKRYDLDAVGLKNMAABDMAEVPIVWQPYOMRQIVTRKILTEFEEEAAPFLPGAGAPRYLYLASNSNKGHRGRYRIQLISFAGRPFLPQESPIEKRAFTMRHIAVTRKKEEPSSSSIFNNDPMTPTVDPEFISNETIAGADLVAATAGTAGVLAHTHAADIPRTVYTAGNSVGFELRPYNPFDDPFSHADSIFREGGDATACEVNPPLACLSTQATCAPEIPIAFSHSGRAYRDN"
	misc_feature
	166..180
	/note="intracellular domain; unclassified site"
	misc_feature
	181..243
	/note="transmembrane-region site"
	misc_feature
	244..2460
	/note="extracellular domain; unclassified site"
	misc_feature
	1576..1578
	/note="topaguinone cofactor binding site; unclassified site"
BASE COUNT	940 a 1191 c 1039 g 1040 t
ORIGIN	
Query Match	42.2%; Score 1705.6; DB 10; Length 4210;
Best Local Similarity	84.1%; Pred. No. 0;
Matches 1924; Conservative	0; Mismatches 364; Indels 0; Gaps 0;

D	b	165	AATGACCCAGGAAGACACCCATTAGTGTCTCTCTGGCTCTGGCTGTATCCACTCTTTGGCTTT	224
Q	y	220	GGTTTGTGTCCTGCTGGTGGGCAGGGGGTGGAGATGGGGGTGAACCCAGCCAGCTTCCCA	279
D	b	225	GGTTTGTGTCCTGCTAGCTAGCTGCAGAGAGGAGATGGGGGTGAGCTAGGCCACTCTTCA	284
Q	y	280	TTGGCCCTCTGTATCTCCCAATGGCCCAAGCCTTGGACACACCCTTGGCCAGAGCCACTGTT	339
D	b	285	TTGGCCCTCTGTCTTCTCTAATGTGCAGCCCCGGGACACACCCTTAGCCAGAGCCAGCCGCTT	344
Q	y	340	TGCAGACCTGAGCCAGAGAGAGCTGAGAGGCGTGTATCGCTTTCTGTACCCAGCGGCTGGG	399
D	b	345	TGCAGACCTGAGCCAGAGAGAGCTGAGAGCTGTATAGCTTCTGTACCAAGACCTGGG	404
Q	y	400	GCCAGGCTGGTGATGTCAGAGCCCAAGGCCCGCCCTCGAGACACTGTCTTCTCACTGGA	459
D	b	405	GCCAGGCTGGTGATGTCAGAGCCCAAGGCCCGCCCTCGAGACACTGTCTTCTCACTGGA	464
Q	y	460	GTTTGAGCTGGCTCCCAAGGCTGCGAGCCCTGGCTCACTTTGGACAGGGGAGACCCCCACC	519
D	b	465	GTTTGAGCTGGCTCCCAAGGCTGCGAGCCCTGGCTCACTTTGGACAGGGGAGACCCCCACC	524
Q	y	520	TGCCCGGAGAGGACATGAGGCATCTCTTCTTTGGAGGCAACCCAGCCCAAGCTGAGTA	579
D	b	525	TGTGAGAGAGGCAATTGGGCATCATCTTCTTTGGTGGACAAACCCAGCCCAAGCTGAGTA	584
Q	y	580	GCTGTGTGGGGGACATGCTCTCACCCCTCCTACATCGGAGGAGTACTGTGGAGGCTGA	639
D	b	585	GCTGTGTGGGGGACCCCTCGCTCACCCCTCCTACATCGGAGGAGTACTGTGGAGGCTGA	644
Q	y	640	TGGAGGCCCCCTGGCCCTATACACGACGCCGCCGTGTCTTTCGAAGAGTACCTGGACATGA	699
D	b	645	GCGGCGCCCCCTGGCCCTATTAACGGGCTCCTGTGTGGACAGAGATATACAGATATCA	704
Q	y	700	CCAGATGATCTTAAACAGAGAGCTGGCCCCAGGGCTTGGGCTCTGCACACTGTGCTT	759
D	b	705	GGAAGTATCTTCCACAGAGAGAGCTGGCCCCAGAGCTTCTGAGCTCTCTCAATACAGTGGCTT	764
Q	y	760	CTACAGACACCGGGGAGCGAACTGTGTACAAATGACACAGGCTCCCCGTGTGTGCAATC	819
D	b	765	CTACAAACACACGAGGAGACAACTGTTAACAATGACATGACAGCCCCCGTGTGTGCAATC	824
Q	y	820	AGGGGACCGGGGACACCTGTGTGGCCCTCTACTCAACATCTCGGGGCGTGGGTTCTTCT	879
D	b	825	AGGGGACCGGGGACACCTGTGTGGCCCTCTACTCAACATCTCGGGGCGTGGGTTCTTCT	884
Q	y	880	GCACACGCTGGGCTTGGAGCGCTAGTGAACCAACAGGCCCTTGAGACCTCGCCCGCTGGAC	939
D	b	885	TCAACCACTTGGCTTGGAGCTTCTGATAGATCAATTAAGGCCCTGGAGATCGCCCTGTGGAC	944
Q	y	940	TATCCAGAAAGTGTCTTATCAAGGCCCTACTACGACAGCGCTGGCCACTGGAGGCCCA	999
D	b	945	CATCCAGAAAGTATTCATCAAGGCCCTTACTATGAGAGCTTCACTTCAGCTGGAGAGCA	1004
Q	y	1000	GTTTTAGGCGCGGCTGGTGAATGTGTGTGTATCCAGAAATGCGACAGAGTGGGCTCG	1055
D	b	1005	GTTTTAGGCGCTGGCTGGTGAATGTGTGTGTATCCCAAAATGTATGACAGTGGGCTCG	1060
Q	y	1060	GTCCCTGAAGTCCCTGTGGCCCCCGGGTCCAGCTCCCCCTCTACAGTTTATCCCAAG	1115
D	b	1065	GTCTCTAAAGTCCCAAGTGGCACCGGGGCCAGCTCCCCCTCTGCAAGTTTCAATCCCAAG	1120
Q	y	1120	CCCCCGTCTAGTGTCCAGAGGAAGTGGAGTGGGCTCTCACTAGTGGACTTCTCTTTGG	1175
D	b	1125	ACCCCGTCTAGTGTCCAGAGGAAGTGGAGTGGGCTCTCTCTCTGTGGGCTTTTCTTTGG	1180
Q	y	1180	CCTGAGACATTCAGTGGCCCAAGGATCTTTGACGTTTCCGCTTCCAAGAGAAAGTACTAGT	1235
D	b	1185	CTTGTGAGCTTCTAGTGGCCCAAGGATCTTTGATATCCGCTTCCAAGGAGAAAGGTTGC	1240
Q	y	1240	TTATGAGATTAAGCTCCAAAGGCTTGGCATCTATGTGTGAATTTCCCAAGCACAAT	1295

Dd	1245	CTATGAAATTCAGTGTCCAGAGAGCCCAATAGCCCTCTACGGTGGAAATTTCCCAAGCATCAAT	1304
Qy	1300	GACGACCCGCTATGATGGATGGAGGCTTTGGCATGGGCAAGTACACAGCCGCTGACCG	1354
Dd	1305	GTCCACACCTGCTACGTGGAGCGGTACCTTTGGCATTTGGCAAACTCTTACCCGCTATCCG	1364
Qy	1360	TGGGATGACCTGCCCCCTACTTTGGCCACCTACGTGGACTGGCACTTCTTTTGGAGTCCCA	1419
Dd	1365	AGGGGTAAACGTGCTCTTACCTGCGCACCTATGTGGAGCTGGCACTTCTTGTAATCTCA	1424
Qy	1420	GGCCCCCAAGCAATACGTGATGCTTTTGTGTGTTGAACAGAACAGGCGTCCCT	1479
Dd	1425	GGCCCCCAAGCACTACGTGTATGCTATTTTGTGTGAACAGAACAGGCGTCCCACT	1484
Qy	1480	GGGGGACACCACTACAGATCTCTCTACGTGCACTTTGGGGGGCTTGGGAAACGCTCT	1539
Dd	1485	CGGGGGGACCACTACAGATTTCTACTCCATTTTGGGGGGTGTGGGAGCGCTCT	1544
Qy	1540	GGTGTGAGATCTATGTGCCACTTGTCAACATGATGATGTGGGATACGTTTCCA	1599
Dd	1545	TGTGTGATGATCCGATCTTACTTGGTCAATTCAGACTACATATGGGAAATGCTTCCA	1604
Qy	1600	CCCCAGTGGGGCCATAGAAATACGATTTCTATGGCACGGGCTACATACGTGGCATTCCT	1659
Dd	1605	CCCCAATGGGGCCATAGAAATTCACAAATTCACGGCACGGCTATACGTCCAGCTTCTT	1664
Qy	1660	CTTTGTGCTACTGGGAAATACGGGAACAAAGTGTCAAGACACACCCGTGGGACGCTCA	1719
Dd	1665	CTTCGCTCTGTGTAGAAATTTGGGAACCAAGTTGGGGCCGACACGCTGGGACGCTACA	1724
Qy	1720	CACCCACAGCCGCCACTTCAAGGTGATCTGATCTTACGAGGACTGTGAGATCGGCTGTG	1779
Dd	1725	CACCCACAGCCCTCACTTCAAAAGTGATCTGATGTGGAGGGCTGGAAGAACTGGGCTTG	1784
Qy	1780	GGCGGAGATATGGTCTTTGTCCCATAGCTGGCTGGCCCTGGAGCCCTGAGCACACAGTCA	1839
Dd	1785	GGCAAGATATTTGGCTTTTGTCCCAAGATTTGACTTGGCAACGGGAGTACCAAGATCA	1844
Qy	1840	GAGGTGCAAGTACCCGGAAGCGCTGGAGATGGAGGAGAGCGCGGCTTCCTGTGGG	1899
Dd	1845	GAGGCTGCAAGTGTACTGGAAAGCTGCTGGAGACAGAGGGAGGAGCTGCTTCCACATGGG	1904
Qy	1900	AAGCGCACCCCTGCTACCTGTACCTGTGGCGACACAACACACAAGTGGGGTCAACC	1959
Dd	1905	GGGGCGCACCCCAAGCTTACCTGTACCTGGCGAGTAACACACACAAAGTGGGGTCAATG	1964
Qy	1960	CCGGGGCTACCGCATCCAGATGCTCAAGCTTTGCTGGAGAGCCGCTGCCCAAAACAGCTC	2019
Dd	1965	GGGGGGCTACCGCATCCAGATTACTCAAGCTTTGCTGGAAAGCCCTTGCCCAAGAAAGTCC	2024
Qy	2020	CATGGCAGAGGCTTCAAGCTGGGGAAGATACAGCTGGGTGTGACCCAGCGGAAGAGGA	2079
Dd	2025	CATGAGAAAGCTTCAAGCTGGGGGAGGATACACTTGGCTGTGACCCAGAGAAAGGAGGA	2084
Qy	2080	GGAGCGAGTACAGCAGCGCTTTTCAATAGAAATGACCTTTGGGCGCCCACTGTGGATTT	2139
Dd	2085	GGAGCGTACAGCGCTTACGATCTTCAACAGAAACGACCCGTTGAGACCCCACTGTGGACTT	2144
Qy	2140	CAGTGACTTATCAACAATAGAACCAATCTGTGGAAAGATTTGTGTGCTGGGTGACAGC	2199
Dd	2145	CACGACCTTATCAGCAATAGAACCAATCTGTGGAGAGCACTTGGTGGTGGAGAGCG	2204
Qy	2200	TGGTTTCTGCATATCCCAATGACAGACATCTCTTAACACAGTACTGTGGGAAACGG	2259
Dd	2205	TGGGTTTTTACACATCCCTCATCAGAAATATTCCCAAACAGGTGACTGCGGGGAATCC	2264
Qy	2260	CGTGGGCTTATCCCTGACACCTCATTAACCTTTTGAAGAAAGCCCTCTCTACTGTGC	2319
Dd	2265	AGTGGGCTTCTCTCGGCGCATTAATCTTTTGAACAGAGACCCCTCTTCCACTTCC	2324
Qy	2320	CGACTCATCTACTCCGAGGGGACCAAGATCTGGGCTCTGCAAGTCAACCCCTTACG	2379
Dd	2325	TGACTTCATATTTCTCGGAGGGCCAGGATCCACAGGCTGTGAGGTAAACCCCTTGGG	238

Tue May 20 13:18:41 2003

us-10-081-408-1.rge

Page 27

Oy	2380	TTCGCTCCCGAGGCTGTGCCTGTGCCCGACATCCTGGCTTCTCCCAAGGGGCTT	2439
Db	2385	TTCGCTCTCCCAAGACTCTCCACCTGTGCCCCGAAATTCTGCTTCTCCATGGGGGCTT	2444
Oy	2440	CTCTACA	2447
Db	2445	TGCTTACA	2452

```
Search completed: May 20, 2003, 06:42:46
Job time : 10509 secs
```

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 02:38:03 ; Search time 826 Seconds

(Without alignments)
11014.616 Million cell updates/sec

Title: US-10-081-408-1

Perfect score: 4040
Sequence: 1 gtccctccacccttagtcc.....ccccaaaaaaaaaaaaaa 4040

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq.101002:*

- 1: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4039.6	100.0	4040	24	ABL61904
2	4039.6	100.0	4040	24	ABL65211
3	2501	61.9	2501	20	AA28632
4	1017.8	25.2	2572	24	ABK35481
5	714.4	17.7	1051	24	ABK63566
6	422.2	10.5	2441	24	ABL62445
7	422.2	10.5	2441	24	ABL67675
8	366	9.1	401	21	AAC04285
9	320.4	7.9	360	20	AA23582

10	297	7.4	1545	23	AA569129
11	150.6	3.7	747	24	AB054593
12	73.6	1.8	303	21	AAA43972
13	69.8	1.7	289	22	ABA46592
14	69.8	1.7	289	22	ABA64456
15	69.8	1.7	289	22	ABA31594
16	69.8	1.7	289	22	AAK12915
17	69.8	1.7	289	22	AAK38641
18	69.8	1.7	289	22	AAI19441
19	69.8	1.7	289	22	AAI44633
20	69.8	1.7	289	22	AAI05166
21	69.8	1.7	289	22	AB513710
22	66.2	1.6	467	22	ABA51780
23	66.2	1.6	467	22	ABA21610
24	66.2	1.6	467	22	AAK00089
25	66.2	1.6	467	22	AAK25525
26	66.2	1.6	467	22	AAI10148
27	66.2	1.6	467	22	AAI31398
28	66.2	1.6	467	22	AAI00089
29	66.2	1.6	467	22	AB500093
30	64.8	1.6	416	24	ABN90024
31	60	1.5	60	24	ABN40805
32	60	1.5	60	24	ABN40805
33	58.2	1.4	36998	24	AA235981
34	55.6	1.4	10732	21	AAA10594
35	51.6	1.3	4386	21	AA289047
36	51.6	1.3	4386	21	AA234997
37	51.6	1.3	65921	21	AA289046
38	51.4	1.3	19243	22	AAK71662
39	51.2	1.3	2035	22	AAH17070
40	49.4	1.2	51	22	AAI29253
41	49.4	1.2	7104	24	ABL70395
42	49.4	1.2	7104	24	ABL34588
43	49.4	1.2	9707	24	ABL33420
44	48.4	1.2	10028	22	AAK68407
45	48.2	1.2	6146	22	ABA07752

ALIGNMENTS

RESULT 1	ABL61904	standard: DNA; 4040 BP.
XX	ABL61904;	
AC	ABL61904;	
XX		
15-MAY-2002	(first entry)	
DT		
XX		
DE	Colon adenocarcinoma related gene sequence SEQ ID NO:241.	
XX		
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;	
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;	
KW	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;	
XX	gene: ds.	
OS	Homo sapiens.	
XX		
PN	WO200194629-A2.	
XX		
13-DEC-2001.		
PD		
XX		
30-MAY-2001;	2001WO-US10838.	
PF		
XX		
05-JUN-2000;	2000US-209473P.	
PR	05-JUN-2000;	2000US-209531P.
PR	18-SEP-2000;	2000US-233133P.
PR	18-SEP-2000;	2000US-233617P.
PR	20-SEP-2000;	2000US-234009P.
PR	20-SEP-2000;	2000US-234034P.
PR	20-SEP-2000;	2000US-234052P.
PR	22-SEP-2000;	2000US-234509P.
PR	22-SEP-2000;	2000US-234567P.

DNA encoding novel
Human ovarian anti
Human secreted exp
Human breast cell
Human foetal liver
Probe #10060 for g
Human brain expres
Human bone marrow
Probe #13319 used t
Probe #5157 used t
Human genome-deriv
Human foetal liver
Probe #76 for gene
Human brain expres
Human bone marrow
Probe #81 for gene
Probe #84 used to
Probe #80 used to
Human genome-deriv
Mouse clone IMX3-7
Human spliced tran
Human transporter
Gene encoding a su
Human nlr1n DNA.
NBS1 gene associat
Human nlr1n DNA.
Human immune/haema
Human cDNA sequenc
Human SNP oligonuc
Chemically treated
Human metastasis a
Human immune syste
Human immune/haema
Human ovarian and

Qy	1141	AAATCGAGTGGCTCTCTCACTGTGGACTTTCCTCTTTGGCTCGAGCATTCAGTGGCC	1200
Db	1141	AAATCGAGTGGCTCTCTCACTGTGGACTTTCCTCTTTGGCTCGAGCATTCAGTGGCC	1200
Qy	1201	AAAGATTTAGAGTTTGCATCCAAAGGAAAGACTAGTTTATGATGAAACCTTCAGA	1260
Db	1201	AAAGATTTAGAGTTTGCATCCAAAGGAAAGACTAGTTTATGATGAAACCTTCAGA	1260
Qy	1261	GGCTTGGCCATCTATGTGTGGAAATTCCTCCAGCAATAGAGACCGCTATGTGATGG	1320
Db	1261	GGCTTGGCCATCTATGTGTGGAAATTCCTCCAGCAATAGAGACCGCTATGTGATGG	1320
Qy	1321	AGGCTTTGGCATGGGCAAGTACACAGCCGCTGACCCGTGGGGGTGAGTACCTCCCTACTT	1380
Db	1321	AGGCTTTGGCATGGGCAAGTACACAGCCGCTGACCCGTGGGGGTGAGTACCTCCCTACTT	1380
Qy	1381	GGCCACCTACGTGGACTGGCACTTCCTTTTGGAGTCCAGAGCCCCCAAGACAATACGTGA	1440
Db	1381	GGCCACCTACGTGGACTGGCACTTCCTTTTGGAGTCCAGAGCCCCCAAGACAATACGTGA	1440
Qy	1441	TGCTTTTGTGTGTTTGAACAGAACCGAGGCTCCCTCGGGGGCAACACTAGATCT	1500
Db	1441	TGCTTTTGTGTGTTTGAACAGAACCGAGGCTCCCTCGGGGGCAACACTAGATCT	1500
Qy	1501	CTATCCGCATCTACTTTGGGGGCTTTCGGGAAAGCGGTGTCGTCACATCTATGTGCAC	1560
Db	1501	CTATCCGCATCTACTTTGGGGGCTTTCGGGAAAGCGGTGTCGTCACATCTATGTGCAC	1560
Qy	1561	CTTGCTCACTATGACTATGTGTGGATACGGTCTTCCACCCCACTGGGGCCATAGAAAT	1620
Db	1561	CTTGCTCACTATGACTATGTGTGGATACGGTCTTCCACCCCACTGGGGCCATAGAAAT	1620
Qy	1621	ACGATTTATGCCACGGGGTACATACAGTCGCGCATTCCTCTTTGGTGTACTGGGAAGTA	1680
Db	1621	ACGATTTATGCCACGGGGTACATACAGTCGCGCATTCCTCTTTGGTGTACTGGGAAGTA	1680
Qy	1681	CGGGAAACCAAGTGTACAGAGCACACCTTGGGCAAGGTCCACACCCACAGCCCACTTCAA	1740
Db	1681	CGGGAAACCAAGTGTACAGAGCACACCTTGGGCAAGGTCCACACCCACAGCCCACTTCAA	1740
Qy	1741	GGTGGATCTGGATGTACAGAGACTGGAGAACTGGGCTGGGCGGAGATATGCTTTGCT	1800
Db	1741	GGTGGATCTGGATGTACAGAGACTGGAGAACTGGGCTGGGCGGAGATATGCTTTGCT	1800
Qy	1801	CCCCATGGCTGTGCCCTGAGCCCTTAGACACAGCTGCAGAGCTGCAGAGTACCAGGAA	1860
Db	1801	CCCCATGGCTGTGCCCTGAGCCCTTAGACACAGCTGCAGAGCTGCAGAGTACCAGGAA	1860
Qy	1861	GCTGCTGGAGATGGAGAGCAGGCGCCTTCTCGTGGAGAGCGCACCCCTCGCTACT	1920
Db	1861	GCTGCTGGAGATGGAGAGCAGGCGCCTTCTCGTGGAGAGCGCACCCCTCGCTACT	1920
Qy	1921	GTACTTGGCCAGAACCAACAGCAAGTGGGGTCAACCCCGGGGCTACCGCATTCAGAT	1980
Db	1921	GTACTTGGCCAGAACCAACAGCAAGTGGGGTCAACCCCGGGGCTACCGCATTCAGAT	1980
Qy	1981	GCTCAGCTTGTCTGGAGAGCCGCTGCCCAAAAACAGCTTCATGCGAGAGGCTTCAGCTG	2040
Db	1981	GCTCAGCTTGTCTGGAGAGCCGCTGCCCAAAAACAGCTTCATGCGAGAGGCTTCAGCTG	2040
Qy	2041	GGAAAGGTACCACTGGCTGTACCCAGCGGAAAGAGAGAGGCCCACTAGCAGCAGCT	2100
Db	2041	GGAAAGGTACCACTGGCTGTACCCAGCGGAAAGAGAGAGGCCCACTAGCAGCAGCT	2100
Qy	2101	TTTCAATCAGATATACCTTTGGGCCCCCACTGTGATTCATCAGCTTCATCAACAATGA	2160
Db	2101	TTTCAATCAGATATACCTTTGGGCCCCCACTGTGATTCATCAGCTTCATCAACAATGA	2160
Qy	2161	GACCATTTGCTGGAAAGATTTGGTGGCTGTGGGTAGACAGCTGGTTTTCTGTGCATATCCACA	2220
Db	2161	GACCATTTGCTGGAAAGATTTGGTGGCTGTGGGTAGACAGCTGGTTTTCTGTGCATATCCACA	2220

[illegible]

D	3301	TTTGGGGGTCGATGATTAATGAAGAAATGGCCATTTTGTATCCAGGGGCTGTGATTTCTGAA	3360
Q	3361	CAGCCTTTTGGGAGGCTGGAGTGGAAACAAAGGTTGGGCAATCAAAGTGAAGACCAA	3420
D	3361	CAGCTGTGTTGGGAGGCTGAGAGGAAACAAAGGTTGGGCAACAAAGATGAAAGCCAA	3420
Q	3421	GGCCCTCAACACTCAGACCAACCCAGGAGGAGGCTGTCCATCACTATTCAGGCAATGGGA	3480
D	3421	GGCCCTCAACACTCAGACCAACCCAGGAGGAGGCTGTCCATCACTATTCAGGCAATGGGA	3480
Q	3481	ATGAGCTGGGGCCCTGGGTGAGGTGGGGGCTGTGGCCTAGTGGGGAGGGGCTGGCTGGGT	3540
D	3481	ATGAGCTGGGGCCCTGGGTGAGGTGGGGGCTGTGGCCTAGTGGGGAGGGGCTGGCTGGGT	3540
Q	3541	GGGGGAGGGGCTGGCTGGGTCCAGGCTTGGGCTCCATTTCCATCACTGTGTCCCTCTG	3600
D	3541	GGGGGAGGGGCTGGCTGGGTCCAGGCTTGGGCTCCATTTCCATCACTGTGTCCCTCTG	3600
Q	3601	AGCTGTGATTTGGGAGTGGGGACAAAGAAATAGCAAGAGATGAGAAACAGAAATTTT	3660
D	3601	AGCTGTGATTTGGGAGTGGGGACAAAGAAATAGCAAGAGATGAGAAACAGAAATTTT	3660
Q	3661	TTTCTCTAAAGGACTGGTTAAATCAATTCGTATACAGCCTTACAAATACATAGATGTCAG	3720
D	3661	TTTCTCTAAAGGACTGGTTAAATCAATTCGTATACAGCCTTACAAATAGATGTCAG	3720
Q	3721	CTAAAAAATAATGTATGTATCTTATATACATATATATATATCTTCAGGTCGAAAAAGC	3780
D	3721	CTAAAAAATAATGTATGTATCTTATATACATATATATATATCTTCAGGTCGAAAAAGC	3780
Q	3781	AAGCCACAGAAATGTGTATATGACGCACTTCCCATTTGTCTTCAGAAAGAGTAGAATATA	3840
D	3781	AAGCCACAGAAATGTGTATATGATGACGCACTTCCCATTTGTCTTCAGAAAGAGTAGAATATA	3840
Q	3841	AACACATATATGCTTATGTATGCTCATTCAGAAATAATGGGTAAACACTGATTACTTTGG	3900
D	3841	AACACATATATGCTTATGTATGCTCATTCAGAAATAATGGGTAAACACTGATTACTTTGG	3900
Q	3901	GAGGGGAACAGTAGTGTGAGGACAGAGAGGAGGAGGCTTAAACACTTACACCTTTTG	3960
D	3901	GAGGGGAACAGTAGTGTGAGGACAGAGAGGAGGAGGCTTAAACACTTACACCTTTTG	3960
Q	3961	TACATTTTGAATTTTGAACCATGTGACTGATTAATCTATTCAAATTAACATTAATATGGG	4020
D	3961	TACATTTTGAATTTTGAACCATGTGACTGATTAATCTATTCAAATTAACATTAATATGGG	4020
Q	4021	CCCCAAAAAATTTTTTTTTTTTAAAAA 4040	
D	4021	CCCCAAAAAATTTTTTTTTTTTAAAAA 4040	

RESULT 2	
ABL65211	
ID	ABL65211 standard; DNA; 4040 BP.
XX	
AC	ABL65211;
XX	
DT	15-MAY-2002 (first entry)
XX	
DE	Lung cancer related gene sequence SPO ID NO:3548.
XX	
KM	Human; Cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KM	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
gene; ds.	
XX	
OS	Homo sapiens.
XX	
PN	WO200194629-A2.
XX	
PD	13-DEC-2001.
XX	

XX PF 30-MAY-2001; 2001MO-USJ0838.
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235740P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237315P.
PR 02-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.

(AVAL-) AVALON PHARM.
XX PA
XX PA
PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
PI Soppe DR, Weaver Z;
XX
XX WPI; 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a
chemical agent to be tested for anti-neoplastic activity, and
determining a change in expression of a gene of a signature gene set -

Claim 1; SEQ ID 3548; 44pp; English.

The present invention describes a method (M1) for screening for an
anti-neoplastic agent. The method involves exposing cells to a chemical
agent to be tested for anti-neoplastic activity, determining a change in
expression of at least one gene (I) of a signature gene set, where (1)
comprises a sequence (S) selected from 8447 sequences (given in ABU61664
to ABU70110), or is at least 95% identical to (S), where a change in
expression is indicative of anti-neoplastic activity. (I) has cyclostic
activity and can be used in gene therapy. M1 can be used for screening
an anti-neoplastic agent, and can be used for producing a product which
is the data collected with respect to the anti-neoplastic agent as a
result of M1, and the data is sufficient to convey the chemical
structure and/or properties of the agent. M1 can be used in the
treatment of cancer such as colon, breast, stomach, lung, thyroid,
adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,

CC Infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilm's tumour.

XX Sequence 4040 BP, 825 A; 1183 C; 1097 G; 934 T; 1 other:

Query Match 100.0%; Score 4039, 6; DB 24; Length 4040;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTCCACCCCTAGTATCCAGGACATGACTACCGGGAACTCAGCCAGAGTCCGGA 60
 DB 1 GTCTTCCACCCCTAGTATCCAGGACATGACTACCGGGAACTCAGCCAGAGTCCGGA 60
 QY 61 GGGCCCGACCCCGCTCCAGAGGCAACAGAGCCCGCTTGGCTGGGTGAGATACATTG 120
 DB 61 GGGCCCGACCCCGCTCCAGAGGCAACAGAGCCCGCTTGGCTGGGTGAGATACATTG 120
 QY 121 CTCTCTTGGTGAATCAAGCTGTCCCTCTTCTGGGAAATGACAGAAAGCAATCT 180
 DB 121 CTCTCTTGGTGAATCAAGCTGTCCCTCTTCTGGGAAATGACAGAAAGCAATCT 180
 QY 181 CGTGTCTCATTTTGGCCGCTATCAACATCTTGGCTTGGTGTCTGCTGTGGTGG 240
 DB 181 CGTGTCTCATTTTGGCCGCTATCAACATCTTGGCTTGGTGTCTGCTGTGGTGG 240
 QY 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTTCCCATGGCCCTCTATCTCCAG 300
 DB 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTTCCCATGGCCCTCTATCTCCAG 300
 QY 301 TGCCACGCTTGGACACACCCCTGGCAGAGCAGCTTTGGACAGCTTGAGCCGAGAGA 360
 DB 301 TGCCACGCTTGGACACACCCCTGGCAGAGCAGCTTTGGACAGCTTGAGCCGAGAGA 360
 QY 361 GCTGACGGCTGTATGGCTTTCTGACCCAGCGGCTGGGGCCAGGGCTGTGATGAGC 420
 DB 361 GCTGACGGCTGTATGGCTTTCTGACCCAGCGGCTGGGGCCAGGGCTGTGATGAGC 420
 QY 421 CGAGGCGCCGCTCGGACAACTGTGTCTCTGAGTGAATTCAGAGTGGCTCCCAAGG 480
 DB 421 CGAGGCGCCGCTCGGACAACTGTGTCTCTGAGTGAATTCAGAGTGGCTCCCAAGG 480
 QY 481 TGCAGCCCTGTGCTACTTGGACAGGGGGAGCCGCCACCTGCCCCGGAGGACATGGCCAT 540
 DB 481 TGCAGCCCTGTGCTACTTGGACAGGGGGAGCCGCCACCTGCCCCGGAGGACATGGCCAT 540
 QY 541 CGTCTTCTTGGAGGCAACCCCAAGCAGTGAAGTGAAGTGTGGGGCCACTGGC 600
 DB 541 CGTCTTCTTGGAGGCAACCCCAAGCAGTGAAGTGAAGTGTGGGGCCACTGGC 600
 QY 601 TCACCCCTCTATATGAGGAGCTGAGTGTGAGCCGTCATGAGAGCCCGCTGCTATCA 660
 DB 601 TCACCCCTCTATATGAGGAGCTGAGTGTGAGCCGTCATGAGAGCCCGCTGCTATCA 660
 QY 661 CCGAGCGCCCGCTGTCTCCAGAGTACTGAGCATAGACAGATGATCTTCAACAGAGA 720
 DB 661 CCGAGCGCCCGCTGTCTCCAGAGTACTGAGCATAGACAGATGATCTTCAACAGAGA 720
 QY 721 GCTGCCCCAGGCTTGGGGTCTCCACACTGTGCTTTACAGACAGCGGGAGCGAA 780
 DB 721 GCTGCCCCAGGCTTGGGGTCTCCACACTGTGCTTTACAGACAGCGGGAGCGAA 780
 QY 781 CCTGTGACAAATGACAGGCTCCCGTGGTGTGCAATGAGGGAGACCGGGCCACTGCT 840
 DB 781 CCTGTGACAAATGACAGGCTCCCGTGGTGTGCAATGAGGGAGACCGGGCCACTGCT 840
 QY 841 TGCGCTCTACTACACATCTCGGGCTGTGGTCTTCTCTGACACAGTGTGGGCTGTGAGCT 900
 DB 841 TGCGCTCTACTACACATCTCGGGCTGTGGTCTTCTCTGACACAGTGTGGGCTGTGAGCT 900
 QY 901 GCTAGTGAACACAGGCGCTTACCCCTGCGCTGGAGCTATCCAGAAAGTGTCTATCA 960
 DB 901 GCTAGTGAACACAGGCGCTTACCCCTGCGCTGGAGCTATCCAGAAAGTGTCTATCA 960

QY 961 AGGCGGCTACTAGACAGCCTGGCCGACGTGGAGGCGCCCAATTTGAGGCGCGCTGTGAA 1020
 DB 961 AGGCGGCTACTAGACAGCCTGGCCGACGTGGAGGCGCCCAATTTGAGGCGCGCTGTGAA 1020
 QY 1021 TGTGTGCTGATCCAGACATGAGCAGAGTGGGTCTGTGCTGCTTAAGTCCCTGTGGC 1080
 DB 1021 TGTGTGCTGATCCAGACATGAGCAGAGTGGGTCTGTGCTGCTTAAGTCCCTGTGGC 1080
 QY 1081 CCGGGGTCAAGCTCCCTCTTACAGTCTATCCCAAGGCGCCCGCTTCAAGTCTCAGGG 1140
 DB 1081 CCGGGGTCAAGCTCCCTCTTACAGTCTATCCCAAGGCGCCCGCTTCAAGTCTCAGGG 1140
 QY 1141 AAGTCAGTGGCTCCCTCACTGAGTGTCTCTGCTGGCTGGAGCTTCACTGAGCC 1200
 DB 1141 AAGTCAGTGGCTCCCTCACTGAGTGTCTCTGCTGGCTGGAGCTTCACTGAGCC 1200
 QY 1201 AAGATCTTTGAGCTTCCAGAGAGAAAGCTAGTTATGATTAAGCCCTCAAGA 1260
 DB 1201 AAGATCTTTGAGCTTCCAGAGAGAAAGCTAGTTATGATTAAGCCCTCAAGA 1260
 QY 1261 GGCCTTGGCATATAGTGGAAATTCGCCAGCAGATGACAGCCGCTATGTGATGG 1320
 DB 1261 GGCCTTGGCATATAGTGGAAATTCGCCAGCAGATGACAGCCGCTATGTGATGG 1320
 QY 1321 AGGCTTGGCATGGGCAAGTACACAGGCGCCCTGACCCGCTGGGCTGGACTGCCCTACTT 1380
 DB 1321 AGGCTTGGCATGGGCAAGTACACAGGCGCCCTGACCCGCTGGGCTGGACTGCCCTACTT 1380
 QY 1381 GGCACCTTACGTGAGTGGACATCTCTTTTGAATGCCAGGCGCCCAAGCAATATGCTGA 1440
 DB 1381 GGCACCTTACGTGAGTGGACATCTCTTTTGAATGCCAGGCGCCCAAGCAATATGCTGA 1440
 QY 1441 TGCCCTTGTGTGTTGAACAGAACAGGCGCTCCCTGTGGCGACACCACTGATCT 1500
 DB 1441 TGCCCTTGTGTGTTGAACAGAACAGGCGCTCCCTGTGGCGACACCACTGATCT 1500
 QY 1501 CTACTGCACTACTTTGGGGTCTTGGGAAAGGCTGTGCTGATGATATGTCAC 1560
 DB 1501 CTACTGCACTACTTTGGGGTCTTGGGAAAGGCTGTGCTGATGATATGTCAC 1560
 QY 1561 CTGCTCACTATGATGTGGGATAGGGTCTTCCACCCAGTGGGGCATAGAAAT 1620
 DB 1561 CTGCTCACTATGATGTGGGATAGGGTCTTCCACCCAGTGGGGCATAGAAAT 1620
 QY 1621 ACGATTTATGCCAGGGGCTATCATGAGCTGGGATTTCTCTTGTGTGCTACTGGGAA 1680
 DB 1621 ACGATTTATGCCAGGGGCTATCATGAGCTGGGATTTCTCTTGTGTGCTACTGGGAA 1680
 QY 1681 CGGGAACCAAGTGTGACAGCAGCACCTGGGACAGGTCACACCCAGAGCGCCACTTCAA 1740
 DB 1681 CGGGAACCAAGTGTGACAGCAGCACCTGGGACAGGTCACACCCAGAGCGCCACTTCAA 1740
 QY 1741 GGTGATCTGATGTAGCAGAGACTGGAAGTGGGTGTGGGCGAGATATGCTTTGT 1800
 DB 1741 GGTGATCTGATGTAGCAGAGACTGGAAGTGGGTGTGGGCGAGATATGCTTTGT 1800
 QY 1801 CCCCATGGCTGTGCTTGGAGCCTGTGACACACTGTGAGAGGCTGTGAGAGGCTCCGAA 1860
 DB 1801 CCCCATGGCTGTGCTTGGAGCCTGTGACACACTGTGAGAGGCTGTGAGAGGCTCCGAA 1860
 QY 1861 GCTGTGGAATGAGAGGAGGCGGCTTCTGTGTGGAAAGGCGCCACTGCTACT 1920
 DB 1861 GCTGTGGAATGAGAGGAGGCGGCTTCTGTGTGGAAAGGCGCCACTGCTACT 1920
 QY 1921 GTACTTGGCCAGCAACACAGCAACAGTGGGGTACACCCCGGGGCTTACCGCATCAGAT 1980
 DB 1921 GTACTTGGCCAGCAACACAGCAACAGTGGGGTACACCCCGGGGCTTACCGCATCAGAT 1980
 QY 1981 GCTCAGCTTTGCTGAGAGCGCGCTGCCCAAAACAGTCCATGCGAGAGGCTTCACTG 2040
 DB 1981 GCTCAGCTTTGCTGAGAGCGCGCTGCCCAAAACAGTCCATGCGAGAGGCTTCACTG 2040
 QY 2041 GGAGAGGTACAGCTGCTGTGACCCAGCGGAAGAGAGGAGGCGAGTACAGCAGCGT 2100

Db 2041 GGAGAGGTACCACTGGCTGTGACCCAGCGGAAAGAGAGAGAGCCCACTGACACACCT 2100
QY 2101 TTTCATACAGATGACCCCTTGGGCCCCACCTGTGATTTCACTTCACTCAACAATGA 2160
Db 2101 TTTCATACAGATGACCCCTTGGGCCCCACCTGTGATTTCACTTCACTCAACAATGA 2160
QY 2161 GACCATTTGCTGGAAAGATTTGGTGGCTGGGTGACAGCTGGTTTTCATATCCACA 2220
Db 2161 GACCATTTGCTGGAAAGATTTGGTGGCTGGGTGACAGCTGGTTTTCATATCCACA 2220
QY 2221 TGCAGAGACATCTCTACACAGTACTGGGGAACGGGTGGGCTTCTCCCGACC 2280
Db 2221 TGCAGAGACATCTCTACACAGTACTGGGGAACGGGTGGGCTTCTCCCGACC 2280
QY 2281 CTATTAATCTTTGACGAGAACCCCTCTCTACTCTGCCACTCATCTACTCCGAGG 2340
Db 2281 CTATTAATCTTTGACGAGAACCCCTCTCTACTCTGCCACTCATCTACTCCGAGG 2340
QY 2341 GGAGCAGGATGCTGGGGGCGGAGGTCACACCCCTAGCTTGGCTGGCCCGGCTGCTGC 2400
Db 2341 GGAGCAGGATGCTGGGGGCGGAGGTCACACCCCTAGCTTGGCTGGCCCGGCTGCTGC 2400
QY 2401 CTGTGCCCCGACCTCCCTGCTCTCTCCACGAGGGCTTCTCTCAACTAGCGGTCT 2460
Db 2401 CTGTGCCCCGACCTCCCTGCTCTCTCCACGAGGGCTTCTCTCAACTAGCGGTCT 2460
QY 2461 GGGATGGGGCATGTGGCCAAAGGCTCCAGGGCCAGGGGTGAGGGATGGGGAGAGCTGG 2520
Db 2461 GGGATGGGGCATGTGGCCAAAGGCTCCAGGGCCAGGGGTGAGGGATGGGGAGAGCTGG 2520
QY 2521 GCATGGGGCGGACGCTGTGCTTCCCTTTCTCTGCTGCCAGACCTCTTTTCTCACATAC 2580
Db 2521 GCATGGGGCGGACGCTGTGCTTCCCTTTCTCTGCTGCCAGACCTCTTTTCTCACATAC 2580
QY 2581 CCTCCCTCGATCCGCTCTGAGCCAGAGAGCTCTGATGCTGATGCTGACACAGAG 2640
Db 2581 CCTCCCTCGATCCGCTCTGAGCCAGAGAGCTCTGATGCTGATGCTGACACAGAG 2640
QY 2641 GACACTGAACCTTGTATGACCACTGACTGAGTTCTCATCACAAGAGGCGAGGATGG 2700
Db 2641 GACACTGAACCTTGTATGACCACTGACTGAGTTCTCATCACAAGAGGCGAGGATGG 2700
QY 2701 CCCAGCTGGAGCCGTGGCGGAGGGCTTCCCTAATAGTTCCCTTGTGTCTGTGCT 2760
Db 2701 CCCAGCTGGAGCCGTGGCGGAGGGCTTCCCTAATAGTTCCCTTGTGTCTGTGCT 2760
QY 2761 TTCCGATCTTTTAGGCACTCTCAAGGACTTAAAGGGGCTATTCCCTGGAGAC 2820
Db 2761 TTCCGATCTTTTAGGCACTCTCAAGGACTTAAAGGGGCTATTCCCTGGAGAC 2820
QY 2821 CCAGAGTAGGGTTGCCAGTCTGCAAGTCCATAGCTGAGTGAAGAGATCTTCTGCTC 2880
Db 2821 CCAGAGTAGGGTTGCCAGTCTGCAAGTCCATAGCTGAGTGAAGAGATCTTCTGCTC 2880
QY 2881 ACATTCCTCTCAACGAGTCTTCTCTGCTGCTTCCCTCTCTCTCTCTCTCTCTCTCT 2940
Db 2881 ACATTCCTCTCAACGAGTCTTCTCTGCTGCTTCCCTCTCTCTCTCTCTCTCTCTCT 2940
QY 2941 TCT 3000
Db 2941 TCT 3000
QY 3001 ATCCCTATGTCACAGCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3060
Db 3001 ATCCCTATGTCACAGCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3060
QY 3061 CCTCTCAGCCCTATGGAAGCTCAAGGTCAGGGAGCCCTTAATCAGATGGGCCATCC 3120
Db 3061 CCTCTCAGCCCTATGGAAGCTCAAGGTCAGGGAGCCCTTAATCAGATGGGCCATCC 3120
QY 3121 TGTGTGTGCTTCCCTTGT 3180
Db 3121 TGTGTGTGCTTCCCTTGT 3180

Db 3121 TGTGTGTGCTTCCCTTGT 3180
QY 3181 CTGGGGCTGGGTGTGTTTCAGAGACGCTGCTTGTGATTTGTGTGCTGCTGCTCATG 3240
Db 3181 CTGGGGCTGGGTGTGTTTCAGAGACGCTGCTTGTGATTTGTGTGCTGCTGCTCATG 3240
QY 3241 CTCTCTATAGAGAGATGGTCTATGCTGACAGCAGGCTCAAGTTAGCATTTCAATGA 3300
Db 3241 CTCTCTATAGAGAGATGGTCTATGCTGACAGCAGGCTCAAGTTAGCATTTCAATGA 3300
QY 3301 TTTGGGGGTCAATGATTAATGAAGAAATGAGCAATTTTGTACAGGGCTCTGATTTCTCAA 3360
Db 3301 TTTGGGGGTCAATGATTAATGAAGAAATGAGCAATTTTGTACAGGGCTCTGATTTCTCAA 3360
QY 3361 CAGCTCTTTGGAGGCTGAGTGAACAAGAGGCTGGCATCAAGATGAGAAAGCCAAA 3420
Db 3361 CAGCTCTTTGGAGGCTGAGTGAACAAGAGGCTGGCATCAAGATGAGAAAGCCAAA 3420
QY 3421 GCGCCCTACAACTCCAGCACCCAGCCAGAGAGGGCTGTCCAAATCAATCAGGATGCGA 3480
Db 3421 GCGCCCTACAACTCCAGCACCCAGCCAGAGAGGGCTGTCCAAATCAATCAGGATGCGA 3480
QY 3481 ATGAGCTGGGCTTGGGTGAGGTGGGGTCTGGCTAGTGGGAGGGGCTTGGCTGGGT 3540
Db 3481 ATGAGCTGGGCTTGGGTGAGGTGGGGTCTGGCTAGTGGGAGGGGCTTGGCTGGGT 3540
QY 3541 GGGGAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3600
Db 3541 GGGGAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3600
QY 3601 AGGCTGAGATTTGGGGATGGGGACAAAGAAATAGCAAGATGAGAAACAGAACTTT 3660
Db 3601 AGGCTGAGATTTGGGGATGGGGACAAAGAAATAGCAAGATGAGAAACAGAACTTT 3660
QY 3661 TTTCTTAAAGGACTGGTTAATCAATTCATGATACAGCTTACAAATAGATATGACAG 3720
Db 3661 TTTCTTAAAGGACTGGTTAATCAATTCATGATACAGCTTACAAATAGATATGACAG 3720
QY 3721 CTAAAAAATTAATTTATGCTCTTTATATATATATATATATATATATATATATATAT 3780
Db 3721 CTAAAAAATTAATTTATGCTCTTTATATATATATATATATATATATATATATATAT 3780
QY 3781 AAGGACAGAAATGATATAGGACGACCTCCATTTGTTTCAAGAGGATAGAAATA 3840
Db 3781 AAGGACAGAAATGATATAGGACGACCTCCATTTGTTTCAAGAGGATAGAAATA 3840
QY 3841 AACACATATATGCTTATGATATGCTTATTCAGAAATTAATGGTAACTGATTTTGG 3900
Db 3841 AACACATATATGCTTATGATATGCTTATTCAGAAATTAATGGTAACTGATTTTGG 3900
QY 3901 GAGGGGAACGATAGGTTGAGAGACAGAGAGGAGGCTTTAACTTACACCTTTTG 3960
Db 3901 GAGGGGAACGATAGGTTGAGAGACAGAGAGGAGGCTTTAACTTACACCTTTTG 3960
QY 3961 TACATTTGAATTTTGAACCTGATGATATATATATATATATATATATATATATAT 4020
Db 3961 TACATTTGAATTTTGAACCTGATGATATATATATATATATATATATATATATAT 4020
QY 4021 CCAAAAAAAAAAAAAAAAAA 4040
Db 4021 CCAAAAAAAAAAAAAAAAAA 4040

RESULT 3
AAK28632
ID AAK28632 standard; DNA; 2501 BP.
XX AAK28632;
AC
XX 21-JUN-1999 (first entry)
DT
XX Nucleotide sequence of the vascular adhesion protein-1.
DE
XX

Human: vascular adhesion protein-1; VAP-1; endothelial cell;
lymphocyte; inhibition; amine oxidase; ss.

Homo sapiens.

Key
CDS

Location/Qualifiers

81..2372
/tag- a
/product- "VAP-1"

81..155
/tag- b
/note- "N-terminal tryptic peptide"

93..161
/tag- c
/note- "transmembrane domain"

207..209
/tag- d
/note- "potential O-glycosylation site"

219..221
/tag- e
/note- "potential O-glycosylation site"

215..217
/tag- f
/note- "potential O-glycosylation site"

489..491
/tag- g
/note- "potential N-glycosylation site"

774..776
/tag- h
/note- "potential N-glycosylation site"

960..962
/tag- i
/note- "potential N-glycosylation site"

1854..1856
/tag- j
/note- "potential N-glycosylation site"

1932..1934
/tag- k
/note- "potential N-glycosylation site"

2076..2078
/tag- l
/note- "potential N-glycosylation site"

300..314
/tag- m
/note- "potential N-glycosylation site"

588..602
/tag- n
/note- "V8 peptide"

855..866
/tag- o
/note- "V8 peptide"

870..911
/tag- p
/note- "V8 peptide"

1047..1067
/tag- q
/note- "V8 peptide"

1155..1163
/tag- r
/note- "V8 peptide"

1182..1205
/tag- s
/note- "V8 peptide"

1230..1259
/tag- t
/note- "V8 peptide"

1764..1778
/tag- u
/note- "V8 peptide"

1879..1883
/tag- v
/note- "V8 peptide"

1950..1991
/tag- w
/note- "V8 peptide"

Query Match 61.9%; Score 2501; DB 20; Length 2501;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

81 GCCACAGAGCCCCCTGCTGCTGCGGTGAGATATGCTCTTGGTTGAATCAG 140
1 GCCACAGAGCCCCCTGCTGCTGCGGTGAGATATGCTCTTGGTTGAATCAG 60

141 CTGTCCCTCTGCTGGGAAATGAAACAGAAATCTGCTGCTCATTTGCGCG 200
CTGTCCCTCTGCTGGGAAATGAAACAGAAATCTGCTGCTCATTTGCGCG 200

61 CTGTCCCTCTGCTGGGAAATGAAACAGAAATCTGCTGCTCATTTGCGCG 120
CTGTCCCTCTGCTGGGAAATGAAACAGAAATCTGCTGCTCATTTGCGCG 120

201 TCATCACAATCTTGGCTGTTGTTGTCCTGCTGCTGCGGAGGATGAGATGAGG 260
TCATCACAATCTTGGCTGTTGTTGTCCTGCTGCTGCGGAGGATGAGATGAGG 260

121 TCATCACAATCTTGGCTGTTGTTGTCCTGCTGCTGCGGAGGATGAGATGAGG 180
TCATCACAATCTTGGCTGTTGTTGTCCTGCTGCTGCGGAGGATGAGATGAGG 180

261 AACCCAGCAGCTTCCCAATGCCCCCTGCTGATCTCCAGTGGCCAGCCTTGACACAC 320
AACCCAGCAGCTTCCCAATGCCCCCTGCTGATCTCCAGTGGCCAGCCTTGACACAC 320

181 AACCCAGCAGCTTCCCAATGCCCCCTGCTGATCTCCAGTGGCCAGCCTTGACACAC 240
AACCCAGCAGCTTCCCAATGCCCCCTGCTGATCTCCAGTGGCCAGCCTTGACACAC 240

321 CTGGCCAGAGCAGCTGTTGACAGCTGAGCCGAGAGAGAGCTGAGAGGCTGATGCGCT 380
CTGGCCAGAGCAGCTGTTGACAGCTGAGCCGAGAGAGAGCTGAGAGGCTGATGCGCT 380

241 CTGGCCAGAGCAGCTGTTGACAGCTGAGCCGAGAGAGAGCTGAGAGGCTGATGCGCT 300
CTGGCCAGAGCAGCTGTTGACAGCTGAGCCGAGAGAGAGCTGAGAGGCTGATGCGCT 300

381 TTCTGACCAAGCGCTGGGCGCAGGCGTGTGATGACGCCAGGCCCGCTCGGACA 440
TTCTGACCAAGCGCTGGGCGCAGGCGTGTGATGACGCCAGGCCCGCTCGGACA 440

301 TTCTGACCAAGCGCTGGGCGCAGGCGTGTGATGACGCCAGGCCCGCTCGGACA 360
TTCTGACCAAGCGCTGGGCGCAGGCGTGTGATGACGCCAGGCCCGCTCGGACA 360

441 ACTGTGCTTCTTCTAGTGAATTCAGCTGCTCCCAAGGCTGAGCCCTGAGCTCACTTG 500
ACTGTGCTTCTTCTAGTGAATTCAGCTGCTCCCAAGGCTGAGCCCTGAGCTCACTTG 500

361 ACTGTGCTTCTTCTAGTGAATTCAGCTGCTCCCAAGGCTGAGCCCTGAGCTCACTTG 420
ACTGTGCTTCTTCTAGTGAATTCAGCTGCTCCCAAGGCTGAGCCCTGAGCTCACTTG 420

501 ACAGGGGAGAGCCCCCAGCTGCCCCGAGAGCACTGGCCATGCTCTTTGGAGGCAAC 560
ACAGGGGAGAGCCCCCAGCTGCCCCGAGAGCACTGGCCATGCTCTTTGGAGGCAAC 560

421 ACAGGGGAGAGCCCCCAGCTGCCCCGAGAGCACTGGCCATGCTCTTTGGAGGCAAC 480
ACAGGGGAGAGCCCCCAGCTGCCCCGAGAGCACTGGCCATGCTCTTTGGAGGCAAC 480

561 CCCAGCCCAACGTAGTGAATTCAGCTGCTGAGGCGCTGCTCAACCCCTCTACATGCGGG 620
CCCAGCCCAACGTAGTGAATTCAGCTGCTGAGGCGCTGCTCAACCCCTCTACATGCGGG 620

481 CCCAGCCCAACGTAGTGAATTCAGCTGCTGAGGCGCTGCTCAACCCCTCTACATGCGGG 540
CCCAGCCCAACGTAGTGAATTCAGCTGCTGAGGCGCTGCTCAACCCCTCTACATGCGGG 540

QY 621 ACCTGACTGTGAGACGTCATGAGAGCCCGCTGCCCTATCAACGACGCCCGCTGCTTCC 680
DB 541 ACAGGACTGTGAGACGTCATGAGAGCCCGCTGCCCTATCAACGACGCCCGCTGCTTCC 600
QY 681 AAGAGTACCTGAGACATAGACAGATGATCTTCAACAGAGAGCTGCCAGGCTCTGGGC 740
DB 601 AAGAGTACCTGAGACATAGACAGATGATCTTCAACAGAGAGCTGCCAGGCTCTGGGC 660
QY 741 TTCTCCACACTGTGCTTCTACAGCAGCGGGAGCGAACTGTGTACAAATGACACAGG 800
DB 661 TTCTCCACACTGTGCTTCTACAGCAGCGGGAGCGAACTGTGTACAAATGACACAGG 720
QY 801 CTGCCCCGTGCTGTCATGACAGGGAGCGGGACACTGTTGGCTCTACTACAAATCT 860
DB 721 CTGCCCCGTGCTGTCATGACAGGGAGCGGGACACTGTTGGCTCTACTACAAATCT 780
QY 861 CGGCGCTGGGTTCTTCTGTCACACAGCTGGGCTGAGAGCTGCTAGTAAACACAGAGGCC 920
DB 781 CGGCGCTGGGTTCTTCTGTCACACAGCTGGGCTGAGAGCTGCTAGTAAACACAGAGGCC 840
QY 921 TTGACCTGCGCGCTGAGCTATCCAGAGGTGTTCTATCAAGCGCGCTACTACAGACGCC 980
DB 841 TTGACCTGCGCGCTGAGCTATCCAGAGGTGTTCTATCAAGCGCGCTACTACAGACGCC 900
QY 981 TGGCCCAAGCTGAGAGGCCAGTTTGAAGCGCGCTGCTGTAATGTGTGCTGATCCAGACA 1040
DB 901 TGGCCCAAGCTGAGAGGCCAGTTTGAAGCGCGCTGCTGTAATGTGTGCTGATCCAGACA 960
QY 1041 ATGGCAAGAGGAGGTCCTGGGTCCTGAAGTCCCTGTGCCCGGGGTCACACTCCCTC 1100
DB 961 ATGGCAAGAGGAGGTCCTGGGTCCTGAAGTCCCTGTGCCCGGGGTCACACTCCCTC 1020
QY 1101 TACAGTTCTATCCCAAGAGCGCGCTTCACTGTCAGAGAGTGCAGTGGCTCTCTAC 1160
DB 1021 TACAGTTCTATCCCAAGAGCGCGCGCTTCACTGTCAGAGAGTGCAGTGGCTCTCTAC 1080
QY 1161 TGTGAGATTTCTCTTGTGGCTGAGAGCTTCAATGGCCCAAGATCTTTACGTTGCT 1220
DB 1081 TGTGAGATTTCTCTTGTGGCTGAGAGCTTCAATGGCCCAAGATCTTTACGTTGCT 1140
QY 1221 TCCAGAGAGAGAGCTATGATTAAGCTTCAAGAGAGCTTGGCCATCTATGAGTG 1280
DB 1141 TCCAGAGAGAGAGCTATGATTAAGCTTCAAGAGAGCTTGGCCATCTATGAGTG 1200
QY 1281 GAAATTTCCCAAGCAATGACAGCCGCTATGTGATGAGAGCTTGGCATGGGCAAGT 1340
DB 1201 GAAATTTCCCAAGCAATGACAGCCGCTATGTGATGAGAGCTTGGCATGGGCAAGT 1260
QY 1341 ACACCAAGCGCGCTGAGCGGAGTGAAGTCCCTACTTGGCCCACTACGTGAGTGGC 1400
DB 1261 ACACCAAGCGCGCTGAGCGGAGTGAAGTCCCTACTTGGCCCACTACGTGAGTGGC 1320
QY 1401 ACTTCTTTTGAAGTCCAGAGCCGCCCAAGACATACGTGATGCTTTTGTGTTGAAC 1460
DB 1321 ACTTCTTTTGAAGTCCAGAGCCGCCCAAGACATACGTGATGCTTTTGTGTTGAAC 1380
QY 1461 AGAACAAGGCGCTCCCTGCGGAGACACACATCTCTACTCGCATCTTGGGG 1520
DB 1381 AGAACAAGGCGCTCCCTGCGGAGACACACATCTCTACTCGCATCTTGGGG 1440
QY 1521 GTCTTGGGAGAAAGGCTGCTGCTCAGATATATGTCACCTTGTCAATGACTATG 1580
DB 1441 GTCTTGGGAGAAAGGCTGCTGCTCAGATATATGTCACCTTGTCAATGACTATG 1500
QY 1581 TGTGGATATAGCTTCTTCCACCCAGTGGGCTAGAAATACGATTCATGTCACAGGCT 1640
DB 1501 TGTGGATATAGCTTCTTCCACCCAGTGGGCTAGAAATACGATTCATGTCACAGGCT 1560
QY 1641 ACATCAGCTCGGCACTCTTGTGTGCTACTGGAATACGGAACCAAGTGTGAGAGC 1700
DB 1561 ACATCAGCTCGGCACTCTTGTGTGCTACTGGAATACGGAACCAAGTGTGAGAGC 1620

QY 1701 ACACCTGGGAGCGGTCACACCCACAGCGGCCACTTCAAGGTGATCTGGATTTACAG 1760
DB 1621 ACACCTGGGAGCGGTCACACCCACAGCGGCCACTTCAAGGTGATCTGGATTTACAG 1680
QY 1761 GACTGAGAACTGGGCTGGGCGCAGAGATATAGCTTGTGCCATGAGTGGCCCTGGA 1820
DB 1681 GACTGAGAACTGGGCTGGGCGCAGAGATATAGCTTGTGCCATGAGTGGCCCTGGA 1740
QY 1821 GCCCTGAGCACAGCTGACAGAGCTGACAGGTGACCCGGAAGCTGTGAGATGAGAGC 1880
DB 1741 GCCCTGAGCACAGCTGACAGAGCTGACAGGTGACCCGGAAGCTGTGAGATGAGAGC 1800
QY 1881 AGGCGCGCTTCTGCTGGGAAAGCGCCACCCCTGCTACCTGTACTGCTCCAGAACCA 1940
DB 1801 AGGCGCGCTTCTGCTGGGAAAGCGCCACCCCTGCTACCTGTACTGCTCCAGAACCA 1860
QY 1941 GCAACAAGTGGGCTCACCCCGGGGCTACCGCAATCCAGATGCTCAGCTTCTGAGAGC 2000
DB 1861 GCAACAAGTGGGCTCACCCCGGGGCTACCGCAATCCAGATGCTCAGCTTCTGAGAGC 1920
QY 2001 CGTGGCCCAAAACAGCTCCATGGCGAGAGGCTTCACTGGGAGAGTACCAGCTGGCTG 2060
DB 1921 CGTGGCCCAAAACAGCTCCATGGCGAGAGGCTTCACTGGGAGAGTACCAGCTGGCTG 1980
QY 2061 TGACCCAGCGGAAGAGAGAGCCAGTAGCAGACAGCTTTTCAATCAGAAATGACCTT 2120
DB 1981 TGACCCAGCGGAAGAGAGAGCCAGTAGCAGACAGCTTTTCAATCAGAAATGACCTT 2040
QY 2121 GGGCCCAAGCTGTGATTTCTGACTTCAATCAATGAGAACCATTTGCTGGAAGATT 2180
DB 2041 GGGCCCAAGCTGTGATTTCTGACTTCAATCAATGAGAACCATTTGCTGGAAGATT 2100
QY 2181 TGTGGCGCTGGGTGACAGCTGGTTTCTGCAATCCAGATGACAGAGACATTTCTTAACA 2240
DB 2101 TGTGGCGCTGGGTGACAGCTGGTTTCTGCAATCCAGATGACAGAGACATTTCTTAACA 2160
QY 2241 CAGTGAAGTGGGGAAGGCGCTGGGCTTCTTCCGCCAGCTTAAATCTTTCACGAAG 2300
DB 2161 CAGTGAAGTGGGGAAGGCGCTGGGCTTCTTCCGCCAGCTTAAATCTTTCACGAAG 2220
QY 2301 ACCCTCTTCTACTGTCGCACTGCATCTACTTCCAGAGGGAGAGATGCTGGGCT 2360
DB 2221 ACCCTCTTCTACTGTCGCACTGCATCTACTTCCAGAGGGAGAGATGCTGGGCT 2280
QY 2361 GCGAGTCAACCCCTTACGCTTGTGCCCCAGGCTGCTGCTGCCCCGAGCTCCCTG 2420
DB 2281 GCGAGTCAACCCCTTACGCTTGTGCCCCAGGCTGCTGCTGCCCCGAGCTCCCTG 2340
QY 2421 CCTTCTCCAGGGGGCTTCTCACAACATGAGGGGCTGAGTGGGCAATGGGCCAA 2480
DB 2341 CCTTCTCCAGGGGGCTTCTCACAACATGAGGGGCTGAGTGGGCAATGGGCCAA 2400
QY 2481 GGGCTCCAGGGCCAGAGGTGTGAGAGATGAGGAGCAGTGGGCACTGGGCCGAGCCTGG 2540
DB 2401 GGGCTCCAGGGCCAGAGGTGTGAGAGATGAGGAGCAGTGGGCACTGGGCCGAGCCTGG 2460
QY 2541 TTCCCTCTTCTGTGCCAGAGCTCTCTTTCTTCCACTACC 2581
DB 2461 TTCCCTCTTCTGTGCCAGAGCTCTCTTTCTTCCACTACC 2501

RESULT 4
ABK35481
ID ABK35481 standard; DNA: 2572 BP.
XX ABK35481;
AC
XX
DT 08-MAY-2002 (first entry)
XX
XX Human endometrial cancer related gene, AOC2.
XX
XX Human; ds. gene; endometrial cancer; differential expression;
KW DNA microarray; protein microarray.


```

Db 1449 CCAATGGGCGCATTTGAAGGCGGGTCCATCCACGGGTTATATCACACACACTTCTG 1508
QY 1661 TTGTGCTACTGGG-----AGTACGGGAACCAAGTGTGAGACACACCTTGGGCGAG 1714
Db 1509 AAAGGGGAGAGAGAGGGGCTCTCTTTGGGAACCGTGTGGGGGAAAGAGTGTCTGGGAACG 1568
QY 1715 GTCCACACACCACAGCGCCCACTTCAAGGTGATCTGATGTAGACGAGCATGAGAGACTGG 1774
Db 1569 GTCCACACACACTGCTTCCACTTCAAGGTGATCTGATGTAGAGGAGGCTGAAAAAAGTGG 1628
QY 1775 GTCTGGGCGGAGATGATGATGTTGTGTGTCCTGAGGCTGTGGCCCTGGAGCCGAGACAG 1834
Db 1629 GTGTAGCTGAAGACGTGTGTTTAAACCTGTGGTGTGCCCCCTGGAAACCCGGAGCATGG 1688
QY 1835 CTGACAGAGCTGACGATGACCGGAAAGCTGTGAGATGAGAGACAGCGCCCTTCTC 1894
Db 1689 CTACAGGCGCCACAGCTGATCTGGAGAGGTCTGGGAAAGAGAGAGACTTATGCTTTTCC 1748
QY 1695 GTGGGAAGCGGCGCTACCTGCTACTGTACTGTGCGCCAGCAACACAGACAGATGGGGT 1954
Db 1749 TTGGGAAGCGCCCTACCTGCTACTGTGCTGCTAGCAACACAGACTTATGCTGGGGT 1808
QY 1955 CACCCCGGGGCTACCGCATCCAGATGCTCAGCTTGTCTGAGAGACCGCTGCCAAG 2014
Db 1809 CACGAGCGCGA----- 1820
QY 2015 AGCTCATGCGGAGAGGCTTACGCTGGAGAGATACAGACTGTGCTGAGCCGAGCGAG 2074
Db 1821 -----TACGAGCTTGTGTGATCCACAGAAAG 1847
QY 2075 GAGAGAGAGCCATAGACAGCGTTTCAATGACACCTTGGGCCCCCACTGTG 2134
Db 1848 GAGAGAGAGTACAGAGAGATGATCATATACAGAAATGACATCTGGACACCCACAGT 1907
QY 2135 GATTTCAGTGCATTCATCAATGATGACATGCTGTGAAAGATTTGGTGGCTGGGG 2194
Db 1908 ACCTTGTGCTGATCATCAATGAAACCTCTTAGAGAGAGATCTGTGCTTGGGCT 1967
QY 2195 ACAGCTGTTTTGTGCATATCCACAGATGACATCTTACACAGATGACTGTGGG 2254
Db 1968 ACACGAGCTTCTGCAATTCCTCCATCCGAGGACATCCAAACAGATGACTGTGGG 2027
QY 2255 AAGGCGTGGGCTTCTTCTCGACCCCTATTAATCTTGTGAGAGAGCCCTTCTTAC 2314
Db 2028 AACGAGATGCTTCTTCTTCTCGACCCCTATTAATCTTGTGAGAGAGCCCTTCTTAC 2087
QY 2315 TCTGCCGATCATCTACTTCTCGAGGGGAGCAGATGCTGGGCTTGGAGGTCAACCC 2374
Db 2088 TCCCTGGCAGTGTGATTTGAGAGAGGCGAGATGCTGGGCTTGGAGCATCAATCT 2147
QY 2375 CTAGCTTCTGCGCCGAGGCTGCTGTGCCCCGACCTTCTGCTTCTCCAGGG 2434
Db 2148 GTGGCTTCTTCTTCTTCTGCGAGCTGCGACCTGTGTCCGCACTTACCCCTTCTTACAC 2207
QY 2435 GGCCTCT 2441
Db 2208 GGCCTCT 2214

```

RESULT 5
 ID ABR63566
 ID ABR63566 standard; cDNA; 1051 BP.

XX
 AC ABR63566;
 DT 18-JUN-2002 (first entry)
 DE Rat sequence differentially expressed in response to a hepatotoxin #1473.
 KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 KW differential expression; centrilobular necrosis; steatosis.
 OS Rattus norvegicus.

```

XX 30-JUL-2001; 2001WO-US23872.
XX 31-JUL-2000; 2000US-222040P.
XX 02-NOV-2000; 2000US-244880P.
XX 11-MAY-2001; 2001US-290023P.
XX 15-MAY-2001; 2001US-290645P.
XX 22-MAY-2001; 2001US-292335P.
XX 06-JUN-2001; 2001US-295798P.
XX 13-JUN-2001; 2001US-297457P.
XX 19-JUN-2001; 2001US-298884P.
XX 09-JUL-2001; 2001US-303459P.
XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
XX WPI; 2002-241625/29.
XX
XX Predicting toxic effects of compounds or the progression of these toxic
XX effects by determining the changes in gene expression in tissues or
XX cells exposed to the toxin and comparing these to gene expression in
XX unexposed tissues or cells.
XX
XX Claim 1; Seq ID No 1473; 239pp; English.
XX
XX The invention relates to methods for predicting toxic effects of
XX compounds or the progression of these toxic effects by determining the
XX global changes in gene expression in tissues or cells exposed to the
XX toxin and comparing these to gene expression in unexposed tissues or
XX cells. Also included are methods of predicting at least one toxic
XX effect of a compound or progression of a toxic effect, preferably the
XX hepatotoxicity of a compound, comprising detecting the level of
XX expression in a tissue or cell sample exposed to the compound of two or
XX more genes listed in the specification, where differential expression of
XX the genes is indicative of at least one toxic effect or progression.
XX The method can also be used to identify an agent which modulates the
XX toxic response and predict cellular pathways that a compound modulates
XX in a cell. The methods utilise a set of at least two probes (on a solid
XX support in kit form), where each of the probes comprises a sequence that
XX specifically hybridises to a gene listed in the specification, a computer
XX system comprising a database containing information identifying the
XX expression level in a tissue or cell sample exposed to a hepatotoxin of a
XX set of genes comprising at least two genes listed in the specification,
XX and a user interface to view the information used to present information,
XX identifying the expression level in a tissue or cell of at least one gene
XX listed in the specification. The method is useful for elucidating global
XX changes in gene expression and for identifying toxicity markers in
XX tissues or cell exposed to a known toxin. The genes may be used as
XX toxicity markers in drug screening and toxicity assays. The genes and
XX gene expression information may be used as diagnostic markers for the
XX prediction or identification of the physiological state of tissue or cell
XX sample that has been exposed to a compound or agent. Hepatotoxicity
XX is characterised by centrilobular necrosis and steatosis. The present
XX sequence is an expressed sequence tag (EST) or cDNA derived from a gene
XX which is differentially expressed in response to a hepatotoxic agent.
XX
XX Sequence 1051 BP; 200 A; 318 C; 291 G; 232 T; 10 other;
XX
XX Query Match 17.7%; Score 714.4; DB 24; Length 1051;
XX Best Local Similarity 82.3%; Pred. No. 1,1e-173;
XX Matches 811; Conservative 5; Mismatches 169; Indels 0; Gaps 0;
XX
XX 141 CTGTCCCTCTTCTGTTGGAAATGAAACAGAGACATCTGCTCTCATTTGGCGG 200
XX 66 CTGTCCACCTTCAGAAACATGACCCAGAGACACCTGATGCTCTCTGCGCGGTG 125
XX 201 TCATCACCATCTTGTGCTGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 260

```

Db 126 TCATCACCATTCTTCCTTGGTTTGGTCTTCTAGTGGCAGAGCGAGATGGGGCA 185
OY 261 AACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAGTCCCAAGCTTGGACACAC 320
Db 186 GACTGAGCCCAACCTTTATGCTCCCTCTCTCTAGCCTCAGCCCAAGACACT 245
OY 321 CTGGCAGAGCGCTTTTGCAGACTGAGCCGAGAGAGAGAGTGAAGGCTGATGCGCT 380
Db 246 CTGGCCAGAGCGAGCGCTTTCAGACCTGAGCCCTGAGAGAGCTGACGTGATGAGCT 305
OY 381 TTCTGACCAGCGCTGGGGCCAGGCTGTGTGATGACAGCCAGCGCCGCTCGGACA 440
Db 306 TTCTGATCAGACCTGGGGCCAGGCTGTGTGATGACAGCCAGCGCTGACCTCGGACA 365
OY 441 ACTGTCTCTCTCATGTGAGTGCAGCTGCTCCCAAGGCTGACGCTGAGCTCACTTGG 500
Db 366 ACTGTCTCTCTCATGTGAGTGCAGCTGCTCCCAAGGCTGACGCTGAGCTGAGCTGAG 425
OY 501 ACAGGGGAGAGCCCGCCACCTGCGCGGAGGAGCACTGGCCATCGTCTTCTTGGCAGGCA 560
Db 426 ACAGAGGG 485
OY 561 CCCAGCCCAAGCTGAGTGCAGTGTGTGTGGGGCCACTGCTTCACTCCCTCTCATGTGCG 620
Db 486 CCAGCCCTATGTGAGCGAGTGTGTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 545
OY 621 AGCTAGCTGTGAGGCTATGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 680
Db 546 AGTGTAGCTGTGAGGCTATGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 605
OY 681 AAGATACCTGAGATAGACAGATGATCTTCAACAGAGAGTGGCCCAAGGCTTCTGGGCG 740
Db 606 GAGATATCAGAGATATTCAGAGAGATGATTTTACAGAGAGAGTGGCCCAAGGCTGCTGTC 665
OY 741 TTCTCACCAGCTGTGTCTTCTACAGACCGGGGAGCGAGGAGCTGTGACATGACCGAG 800
Db 666 TCCTCATCTACTGTGTCTTCTACAGACCGGGGAGCGAGGAGCTGTGACATGACCGAG 725
OY 801 CTCCCGGCTGTGATGAGGGGAGCGGGGAGCGGCTGTGGCTTCTCTCAATCATCT 860
Db 726 CCCCCGCTGTGTGATGAGGGGAGCGGGGAGCGGCTGTGGCTTCTCTCAATCATCT 785
OY 861 CGGGCGCTGTGTCTTCTCTGACACAGTGGGCTTGGAGCTGCTAGTGAACCAAGGCGCC 920
Db 786 CAGGGGCTGTGTCTTCTTACCCCTACCCCATTTGGCTTAGAGCTTGTGATGATCACAAGGCC 845
OY 921 TTGACCTGCGCGCTGTGACTATCCAGAGGCTTCTATCAGAGCGGCTACTAGACAGCC 980
Db 846 TGGATCTGTGCTGTGTGACATCCAGAGGCTTCTATCAGAGCGGCTACTAGAGATG 905
OY 981 TGGCCAGCTGTGAGGGCCGCTTGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1040
Db 906 TGGATCTGTGAGGAGCATGTTTGTGAGGCTGTGGCTGTGATGATGTTTGTGGCTCGGAGA 965
OY 1041 ATGGCAGAGTGTGTCTGTGCTCTGTAAGTCCCTGTGCTGCGGGGCTCCAGCTCCCGCTC 1100
Db 966 ATGGCAGAGTGTGTGTGCTGTGCTGTAAGTCTTCAAGTGTGACAGGCGGAGCTCTCTC 1025
OY 1101 TACAGTTCTATCCCAAGGCGCCCG 1125
Db 1026 TGCARTTTCATCCNGARNGNCCNMG 1050

RESULT 6
ABL62445

ID ABL62445 standard; DNA: 2441 BP.

AC ABL62445;

DF 15-MAY-2002 (first entry)

DE Colon adenocarcinoma related gene sequence SEQ ID NO:782.

XX

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
OS Homo sapiens.
XX WO200194629-A2.
PN 13-DEC-2001.
PD 30-MAY-2001; 2001WO-US10838.
PF 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
PA (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrihan S;
PI Soppet DR, Weaver Z;
PI WPI: 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 782; 44pp; English.
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)

Db 2014 ACCAGACACCCCTGGACACCCCGCTGGTCTTTGAGCATCTTCTTACACACAGAGA 2073
 Qy 2163 CCATTCTCTGMAAGATTGGTGGCTGGTGACACTGTTTTCGATATCCACATG 2222
 Db 2074 ACATTGMAATGAGGACCTGGTGGCTGGTGACACTGTTTTCGATATCCACATG 2133
 Qy 2223 CAGAGACATCTCTTACACAGTACTGTGGGAAAGCGGCTTCTTCTCCGAGCCCT 2282
 Db 2134 CAGAGACATCTCTTACACAGTACTGTGGGAAAGCGGCTTCTTCTCCGAGCCCT 2193
 Qy 2283 ATACTCTTCTTACAGAGACCCCTCTTCTACTCTGCGACTC 2325
 Db 2194 TCACTCTTCTTCCAGAGACCCCTCTCTGCAATCCAGAGACAC 2236
 RESULT 7
 ABL67675
 ID ABL67675 standard; DNA; 2441 BP.
 XX
 AC ABL67675;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Oesophagus cancer related gene sequence SEQ ID NO:6012.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001MC-US10838.
 XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-233133P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236881P.
 PR 29-SEP-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.

PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX
 PA (AVAL-) AVALON PHARM.
 PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrihan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.
 DR
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set
 PS
 PS Claim 1; SEQ ID 6012; 44pp; English.
 XX
 XX The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilm's tumour.
 XX
 XX Sequence 2441 BP; 515 A; 789 C; 684 G; 453 T; 0 other;
 SQ
 Query Match 10.5%; Score 422.2; DB 24; Length 2441;
 Best Local Similarity 52.8%; Pred. No. 4.3e-98;
 Matches 1099; Conservative 0; Mismatches 918; Indels 66; Gaps 6;
 Qy 303 CCCAGCCTTGACACACCCCTGGCCAGCCAGCTGTTGACAGACTGACCCGAGAGAGC 362
 Db 160 CCTCCCGGGGAGTCTGCCAGGAAGCAGGGGTTTTCAGACTTAAGCAACCAAGAGC 219
 Qy 363 TGACGGCTGTGATCGCTTCTGTACCCAGGGGCTGGGGCCAGGGCTGTGATCCAGCC 422
 Db 220 TGAAGGAGTGCACAGATCTCTGTGTCAGAAAGGAGCTGAGGCTCAGCCCTCCAGTA 279
 Qy 423 AGGCGCGCCCTGCGACACAGTGTCTTCTAGTGAGAGTGCACCTGCCCAAGGCTG 482
 Db 280 CCACACACATGCGCCAGAACACCGCTGTTTCATCGAGATCTCTGCCCAAGAGTACC 339
 Qy 483 CAGCCCTGCTCACTTGACAGAGGGGAGCCGCCACCTGCCCGGAGGACATGGCCATCG 542
 Db 340 ATGTGCTGAGATTCTGTGATTAAGTGAAGAGCATCTGTGCGGGAAGCCGTCGCTCA 399
 Qy 543 TCTTCTTTGGCAGGACCCAGCCCAACGATGATGAGCTGTGTGGGGCATGTGCTG 602
 Db 400 TCTTCTTTGGCAGGAGAGCATCCCAATGACAGGTTGCTGTGTGGGGCCCTGTGCGAG 459
 Qy 603 ACCCTCTCATGTCGGGAGCTGACTGTGAGGCG--GTCATGAGAGCCCTGCGCTATC 659
 Db 460 GGGCTGCTACATCGAGCACTGTCCCGCAGGCTGTGGTACCAATCTCTCTGGGCAATGA 519
 Qy 660 ACCGAGCCCGCTGCTGTTCCAGAGTAACTGACATAGACAGATGATCTTCAACAGAG 719
 Db 520 GGCCATCTGCACAGAGATATGCCCTCTTACACACACCTTCGAGAGAACCCACCAAGC 579

```

QY 720 AGTGGCCCAAGGCTTGGGCTTCTCCACACTGTGCTTCTACAGACCGGGAGCGA 779
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 580 CCGGCGATCACTTCTCTCAATATACAGAGGCTTCTCAATTCAGAACTGATGACAT 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 780 ACCTGTGCAATATACAGAGGCTCCCGGTGTGTCAATACAGGGAGCCGGACCTGGT 839
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 640 GCGTGGCTTACAGAGTGTGGCCCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 840 TTGGCTCTACTACAACTCTGAGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 899
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 700 TTATCATACAGCGCTATGTAGAAAGC-----TACTTCTGACCCCACTGGGCTGAGC 753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 900 TGGTAGTAACACAGAGCCCTTGACCTGCGCTGACATTCAGAAAGTGTCTATC 959
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 754 TCTGTGTGATCATAGGAGCAACATGTGTGGCTATGGGCGGTGAGAGAGTGTGTCA 813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 960 AAGGCGCTACTACAGACAGCTGGCCAGCTGAGAGGCCATTTGAGGCGCGCTGTGA 1019
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 814 ACGGGAAGTTCTATGGAGCCCAAGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1020 ATGTGTGTGTATCCAGACA-----ATGGCAAGGTGGT 1055
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 874 ACGTGTGTGTCTGTGAGAGCCCGCTGCGTGGGGCAAGGGGATGACAGACAGAGAGC 933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1056 CCTGTGCTGAGTCCCGCTGTGGCCCGGGTCCAGCTCC----- 1095
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 GCGCCCTCTTCTCCCTCCACAGCCCGCGGGAGCTTCCAGCCCAATCATGTGACG 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1096 -CCCTCTACCTTATATCCCAAGGCCCGCTTCACTGTGAGAGAGTGTGAGGCGCT 1154
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 994 GCGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1155 CCTCATGTGTGACTTCTCTTGTGGCTGTGAGCATGTGTGTGTGTGTGTGTGTGT 1214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1054 ACGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1215 TTGCTTCCAGAGAAAGATATGTATGATATGATTAAGCTTCCAGAGGCTTGGCATCT 1274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1114 TGCACCTTGGGGGAGAGGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1275 ATGTGGAATTCCTCCAGACCAATGACAGCCCGCTATGTGTGTGTGTGTGTGTGT 1334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1174 ATGGAGACACACACCTGACAGATGACAGACAGTGTGTGTGTGTGTGTGTGTGT 1233
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1335 GCAAGTACACACGCTTGTGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1394
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1234 GCAGGTGCTCATGATGATTAAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1293
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1395 ACTGGCACTTCTTGTGAGTCCAGAGCCCGCAAGACAAATACGTGTGTGTGTGT 1454
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1294 AACTTTCCTACTATATGATGCTGATGACCGGTCTATATCCCGAGCCCTCTGCTCT 1353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1455 TTGAACAGAACAGGCGCTCCCGCTGGGGAGACAC-----ACTCAGATCTCTACT 1505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1354 TTGAAATGCCACAGAGGCTGCCCTGTGGCGGCACTTAATTCACACTTTAAAGTGTCT 1413
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1506 GCGACTACTTGTGGGCTTGTGGGAAAGGTGTGTGTGTGTGTGTGTGTGTGTGT 1565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1414 TCACTTCTATGCGGGGTGAAGGGCGAGGTGTGTGTGTGTGTGTGTGTGTGTGT 1473
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1566 TCACTATGACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1474 ACAATTATGATTAATTTGGGACTTATCTTACCCCAAGGGGTGTGTGTGTGTGTGT 1533
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1626 TCTATGCCAGGCGTACATCAGCTGCGATCTCTTGTGTGTGTGTGTGTGTGTGTGT 1685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1534 TGCATGCCACAGCTGCTACGCGACGCTTCTACACCCCGAGGGGCTGCGCCACGGCA 1593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1686 ACAAGTGTCAAGACACCTTGGGACGCTTCCACACCGCCCACTTCAAGGTGTG 1745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1594 CTCGCTGTACACCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1746 ATCTGATGTAGCAGACTGGGAACTGGGTCTGGCCGAGAGATATGGTCTTGTGCCCA 1805
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1654 ACTGTGATGTGGCAGGACCAAGAAAGCTTCCAGACACTGTCAATGAAAGTAAACAA 1713
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1806 TGGCTGTGCTGTGAGCCCTGAGACCAAGCTGAGAGGCTCAGAGTACCCGAGGCTGC 1865
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1714 TCACCAACCCCTGAGACCCCAAGACACCGCTGTGTGTGTGTGTGTGTGTGTGTGT 1773
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1866 TGAAGATGAGAGACAGCGCCCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1925
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1774 ACTCTGTGAGCGCGCAGCGCTTCCGCTTCAAAAGAAAGCTGTCAAGTACGTGTCT 1833
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1926 TGGCAGCAACACAGCAACAAAGTGGGTCAACCCCGGGGTATCCGATCCAGATGTCTCA 1985
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1834 TTACCGAGCCCGCAGAGAACCCCTGGGGCCCAAGCGAGGTATCCGCTGCAATCCACT 1893
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1986 GCTTGTGTGAGAGCCGCTGCGCCCAAAACAGCTCATATGGCAGAGGCTTCACTGTGAGGA 2045
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1894 CCATGTGCGGACAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1953
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2046 GGTACCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1954 GGTACCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2013
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2106 ATCAAAATGAGCTTGTGGCCCGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2162
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2014 ACCAGAACGACCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2073
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2163 CCATTGTGTGGAAGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2222
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2074 ACATTGAAATGAGAGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2133
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2223 CAGAGGACTTCTTAACACATGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2134 CAGAGGACTTCTTAACACACACACACACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2193
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2283 ATACTTCTTGTGAGAAAGCCCTCTCTACTGTGCGGACTC 2325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2194 TCACTTCTTCCAGAGAGACCCCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 8
AAC04285
ID AAC04285 standard; cDNA: 401 BP.
XX AAC04285;
AC 06-OCT-2000 (first entry)
AC 06-OCT-2000 (first entry)
DT 06-OCT-2000 (first entry)
DE Human secreted protein 5' EST, SEQ ID NO: 8360.
DE Human; 5' EST, expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
XX (GIST ) GENSET.
XX Dunas Milne Edwards J, Duclert A, Giordano J;
XX WPI: 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX

```

PS Claim 1; SEQ ID 8360; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs with intact 5' ends and can therefore be
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 401 BP; 69 A; 137 C; 103 G; 91 T; 1 other;

SO Query Match 9.1%; Score 366; DB 21; Length 401;
 Best Local Similarity 99.5%; Pred. No. 4.8e-84;
 Matches 377; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GTCCTCCACACCTTATGCCAGCATCTGACTACCGGAACTCAGCCAGATCCGGA 60
 Db 24 GTCTTCCACACCTTATGCCAGCATCTGACTACCGGAACTCAGCCAGATCCGGA 83
 QY 61 GCCCCACACCCCGTCAGAGACCAAGAGCCCGCTTGTGCTGAGCAATACATTG 120
 Db 84 GCCCCACACCCCGTCAGAGACCAAGAGCCCGCTTGTGCTGAGCAATACATTG 143
 QY 121 CTCCTCTTGGTTAAATCAGTGTCTCTTCTGCGGAAATGAAACAGAAATACCT 180
 Db 144 CTCCTCTTGGTTAAATCAGTGTCTCTTCTGCGGAAATGAAACAGAAATACCT 203
 QY 181 CGTCTCTCTATCTGCGCCGTCATCACCATTGCTTGGTTGCTGCTGCTGCTG 240
 Db 204 CGTCTCTCTATCTGCGCCGTCATCACCATTGCTTGGTTGCTGCTGCTGCTG 263
 QY 241 CAGGGGTGAGATGGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGATCTCCAG 300
 Db 264 CAGGGGTGAGATGGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGATCTCCAG 323
 QY 301 TGCCACGCTTGACACACCTTGCGCAGAGCAGCTGTTGTCAGACCTGAGCCAGAGA 360
 Db 324 TGCCACGCTTGACACACCTTGCGCAGAGCAGCTGTTGTCAGACCTGAGCCAGAGA 382
 QY 361 GCTGACGCTGTGATGCGC 379
 Db 383 GCTGACGCTGTGATGCGC 401

RESULT 9
 AA233582
 ID AA233582 standard; cDNA; 360 BP.
 XX
 AC AA233582;
 DT 08-DEC-1999 (first entry)
 XX
 DE Human breast tumour-associated EST 42.
 XX
 KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
 KW Medicaments; gene therapy; treatment; fat metabolism; ss.
 XX
 OS Homo sapiens.
 XX
 PN DE19813835-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 20-MAR-1998; 98DE-1013835.
 XX

PR 20-MAR-1998; 98DE-1013835.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 PA Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
 PI WPL; 1999-528979/45.
 XX
 DR
 XX
 XX Human nucleic acid sequences and protein products from normal breast
 PT tissue, useful for breast cancer therapy
 CC
 PS Claim 3; 130; 206pp; German.

CC This invention describes novel human nucleic acid sequences from normal
 CC breast tissue which have cytostatic activity. The nucleic acid sequences
 CC can be used to produce and isolate full-length gene sequences. They can
 CC be used to express proteins, which can be used as tools to find an
 CC activity against breast cancer. The sequences can be used in sense or
 CC antisense form. They are especially useful for medicaments for gene
 CC therapy to treat breast cancer and for treating illnesses associated
 CC with fat metabolism. AA233541-233610 represent expressed sequence tags
 CC described in the method of the invention.

XX Sequence 360 BP; 119 A; 63 C; 81 G; 97 T; 0 other;

SO Query Match 7.9%; Score 320.4; DB 20; Length 360;
 Best Local Similarity 99.7%; Pred. No. 2.6e-72;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3707 ACAATAGTATCAGCTAAATAATATGTTATGTTATATGTAATATATCTT 3766
 Db 37 ACAATAGTATCAGCTAAATAATATGTTATGTTATATGTAATATATCTT 96
 QY 3767 CAGGTGAAAAAGCAAGCCACAGAAATGTATAGCGCACTTCCATTGTTTCAGAA 3826
 Db 97 CAGGTGAAAAAGCAAGCCACAGAAATGTATAGCGCACTTCCATTGTTTCAGAA 156
 QY 3827 AGGAGTAAATTAACACATTAATGCTTATGATGCTTATGCAATTAATGGGTACAA 3886
 Db 157 AGGAGTAAATTAACACATTAATGCTTATGATGCTTATGCAATTAATGGGTACAA 216
 QY 3887 CTGATTACTTTGGGAGGGAACAGTAGTTGAGACAGAGAGGAGGCTTTAACA 3946
 Db 217 CTGATTACTTTGGGAGGGAACAGTAGTTGAGACAGAGAGGAGGCTTTAACA 276
 QY 3947 CTTACACCCCTTTTGTACATTTTGAATTTGAACATGTGACTGTATACCTATTAAT 4006
 Db 277 CTTACACCCCTTTTGTACATTTTGAATTTGAACATGTGACTGTATACCTATTAAT 336
 QY 4007 AAACATTAATGGGCCAAAAA 4028
 Db 337 AAACATTAATGGGCCAAAAA 358

RESULT 10
 AAS69129
 ID AAS69129 standard; cDNA; 1545 BP.
 XX
 AC AAS69129;
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #4933.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX

PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG04942.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID NO 4933; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIGO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1545 BP; 372 A; 395 C; 429 G; 349 T; 0 other;
Query Match 7.4%; Score 297; DB 23; Length 1545;
Best Local Similarity 98.4%; Pred. No. 6.7e-66;
Matches 300; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1457 GAACAGAACAGAGGCGCTCCCTGGGGGACACCACTCAGATCTTACTGCACTACTT 1516
DB 931 GAGATGAACCAAGGCGCTCCCTGGGGGACACCACTCAGATCTTACTGCACTACTT 990
QY 1517 GGGGCTTGGCGAAACGGTCTGCTCAGATCTATGTCACCTTGCCTCAACTATGAC 1576
DB 991 GGGGCTTGGCGAAACGGTCTGCTCAGATCTATGTCACCTTGCCTCAACTATGAC 1050
QY 1577 TATGTGTGGATACGCTTCTCCACCCAGTGGGGCATAGAAATCAGATTTATGCCACG 1636
DB 1051 TATGTGTGGATACGCTTCTCCACCCAGTGGGGCATAGAAATCAGATTTATGCCACG 1110
QY 1637 GAGTCATCAGCTCGGCACTCTTTGTGCTACTGGGAAGTAGGGGACCAAGTGCA 1696
DB 1111 GGTATCATCAGCTCGGCACTCTTTGTGCTACTGGGAAGTAGGGGACCAAGTGCA 1170
QY 1697 GAGCAGACCTGGGCGACGGTCCACACCAAGCCCACTTCAGATGGATGTGATGA 1756
DB 1171 GAGCAGACCTGGGCGACGGTCCACACCAAGCCCACTTCAGATGGATGTGATGA 1230
QY 1757 GCAGG 1761
DB 1231 GCAGG 1235
RESULT 11

ABQ54593
ID ABQ54593 standard; CDNA: 747 BP.
XX
AC ABQ54593;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HVBK73 CDNA, SEQ ID NO:473.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; chromosome 7q34-ter;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-147878/19.
DR P-PSDB; ABP41516.
XX
PT Isolated nucleic acid molecules, encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 1; SEQ ID NO 473; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC

Sequence 747 BP; 142 A; 240 C; 205 G; 155 T; 5 other;

Query Match	3.7%	Score 150.6	DB 24	Length 747
Best Local Similarity	55.3%	Pred. No. 2.6e-28		
Matches 316	Conservative 1	Mismatches 245	Indels 9	Gaps 1

1096 CCCTCTACAGTTCTATCCCAAGGCCCCCGCTTCAGTGTC CAGGAGTCGAGTGGCCTC 1155

Db 70 CCCCCGCTTGTCACAGCCCCACGGCCCTCGCTTCAGGCTGAGGGCAACGCTGTGCTCTA 129

QY	1156	CTCAGCTGGACCTTCTCCCTTTGGCCCTGAGACATTCAGTGGCCCAAGATCTTTGAGCT	1213
Db	130	CGCGGGCTGGAGCTTTCCTTCGCGGTGGCGTCTCTCCGGGCTGCAAGTCTCAAGT	189
QY	1216	TCGCTTCCAAAGAGAAAGACTAGTTATTAGATAGACCTCCAAAGAGCCCTTGGCATCTA	1275
Db	190	GCACTTGGGGAGAGGAGCATTCCTATAGAGTACACCGTGCAGAAAGAGCGTGGCCGTGA	249
QY	1276	TGGTGGAAATTTCCCAAGACACATAGACAGCCCGCTATGTGATGGAGGCTTTGGCATGGG	1333
Db	250	TGGAGGACACACACCTCGACAGCATGACAGACCAAGTACCTGATGTGGCTGGGGCCCTGGG	309
QY	1336	CAAGTACACCACGCCCTGACCCCGTGGGGTGGAGTCCCTCTTGGCCACCCATGACTGGA	1395
Db	310	CAGGTCACTCATGAGTTAGTACCCCCCGGACATGCAGCTCCGGAGACCGGCACTTCTCTGA	369
QY	1396	CTGGCACTTCTTTTGGAGTCCAGGCCCCCAAGACATAGTATGCTTTGTGTGT	1455
Db	370	CACCTTCACATACATATGATATGCGATACCCCGGTCACTTATACCCGAGCCCTTGCTCTT	429
QY	1456	TGAACAGAACAGAGGCCCTCCCTCTGGGGGACA-----CCACTCAGATCTCTACTC	1506
Db	430	TGAATATCCCAACAGGGGTGCCCTCTGGGGGCACTTAATCCAACTTAAAGGTGGCT	489
QY	1507	GCACTACTTTGGGGGTCTTCCGGAAGACGTGCTGGTGTGTCAGATCTATATCCACTTGGCT	1566
Db	490	CACCTTCTATAGCGGGGTGAAAGGCCAGAGGTGGGTGGTGGCGACAACCTCAACGTGTGA	549
QY	1567	CAACTATAGCTATGTGTGGATAGCTTCTCCACCCCACTGGGGCCATGAATACGATT	1626
Db	550	CAATATGATTACATTTGGGACCTTATCTCTACCCCAAGGGGGATGAGAGCCAAAGT	609
QY	1627	CTATGCCAGGGCTACATACGTCCGGCATTC	1657
Db	610	GCATGCCACGTGACGTCCAGCCCACTTC	640

7
3
2
1
1
3
3
1
2

AAA43972

XX 8813070

XX 71-8176-3000 / 44-984 884-...

DE

DE Human secreted expressed sequence tag SEQ ID NO:547.
XX
KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antitumoral;
KW antiviral; antidiabetic; antitastmatic; vulnerrary; antiparkinsonian;
KW anticulcer; osteopathic; neuroprotective; nootropic; antiporiatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;

KW tumour; infection; depression; psoriasis; ss

OS Homo sapiens

PN WO200021991-A1

PD 20-APR-2000.

15-OCT-1999; 99WO-US24206.

PR 15-OCT-1998; 98US-0104436.

PA (GEM) GENETICS INST INC.

PI Jacobs K, McCoy JM, Lavallee ER, Collins-Racie LA, Evans C,

XX
XX
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
8

	XX	T
	XX	C
	E	+2
	E	2
		g
		C]
		:::
		:g]
		o
		o
		+

PT expressed sequence tags (sESTs), useful for treating various disorders such as autoimmune infectious and central nervous system disorders

XX
XX
27437
E220
340
00355
F2716
b

XX AAAA312E +C AAAAE02E -00000000+ 000

CC

sequence tags (sESTs), isolated from

the tissues were isolated from. The activities include:

- chemotactic; proliferative; immunomodulatory; haematopoietic;
- chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
- cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
- antiallergic; vulnerary; antitumor; osteoprotective; neuroprotective;
- neurogenic; antiparkinsonian; antipsoriatic; cerebroprotective;
- anticonvulsant; and antidepressant. The sESTs can be used for gene therapy and in vaccines. The sESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the sESTs. Proteins encoded by the sESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAAA5926 to AAA45931 represent linker variants which are generated in the exemplification of the present invention.

50 Sequence 303 BP; 46 A; 106 C; 65 G; 86 T; 0 other;

Query Match 1.8%; Score 73.6; DB 21; Length 303;

Matches 103; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2290 CTTGACGAGACCCCTCTCTACTCTGCCGACTCCATCTACTTCCGAGGGGACCAGGA 23

Db 2 CTTGATGAGGACCCCTCCATCTTCTCCCTGGCAGTGTCTACTTTGAGAAGGCCAGGA 61

2350 TGCTGGGGCTGCGAGGTCAACCCCTAGCTTGCCTGCCCGCCAGGCTGCTGCCCTGTGCCC 24

Db 62 TGCTGGGCTCTGCAGCATCATCCTGTGGCCTGCCTCCCGACCTGGCAGCCTGTGCC 121

QY 2410 CGACCTCCCTGCCTTCTCCACGGGGCTTCT 2441

Db. 122 GGACTTACCCCTTCTCTTACCAAGGCTTCT 153

RESULT 13

ID	ABA46592	standard; DNA; 289 BP
XX		
AC	ABA46592;	

AC ABA46592;

AC ABA46592;

AC ABA46592;

```

XX 01-FEB-2002 (first entry)
DT Human breast cell single exon nucleic acid probe #5287.
XX Human; microarray; single exon probe; gene expression: breast;
KW disease; cancer; ss.
XX Homo sapiens.
XX MO200157271-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00662.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes
XX Claim 4; SEQ ID NO 5287; 327pp + sequence listing; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 289 BP; 37 A; 77 C; 102 G; 73 T; 0 other;
XX
XX Query Match 1.7%; Score 69.8; DB 22; Length 289;
XX Best Local Similarity 52.6%; Pred. No. 1.1e-07;
XX Matches 152; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
XX
XX 1758 CAGACTGTGAGACTGGGTGGGCGGAGATGATGCTTTGTCCCATGGCTGCGCCT 1817
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 289 CAGGACCCAAACAGCTTCAGACACTGACATGAAACCTCAACCAACCCCT 230
XX
XX 1818 GGAAGCCCTGAGCAGCTGAGAGCTGACGACCGGAGAGTGTGAGATGAGG 1877
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 229 GAGAGCCCAAGACACGGGTGTCAGCCAACTGTGAGACACGAGTACTCTGGAGC 170
XX
XX 1878 AGCAGGCGGCTTCTCTGTGGAAAGCGCACCCCTCGTACTCTGTGACCAAC 1937
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 169 GCGAGGCGGCTTCCGCTTCAAAAGAGCTGCTAAGTACTGCTCTTACACGCCCC 110

```

```

XX 1938 ACAGCACAACTGGGGTACCCCGGGGCTACCCGATCCAGATGCTGACTTGTGAG 1997
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 109 AGGAGAACCCCTGGGGCCACAGCGACGTACCCCTCAGATCCACATCCAGCCGACC 50
XX
XX 1998 AGCGGTGCCCCAAACAGCTCCATGGCGAGAGGCTTACCTGGGAGAG 2046
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 49 AGTGTGTCCCGCCAGCTGCGAGGAGAGAGCGCATCCTGAGGCAAG 1
XX
XX RESULT 14
XX ABA64456/C
XX ID ABA64456 standard; DNA; 289 BP.
XX
XX ABA64456;
XX AC
XX AC
XX AC
XX 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #12761.
XX XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX KW Homo sapiens.
XX OS
XX OS
XX PN MO200157277-A2.
XX PD
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX Claim 4; SEQ ID NO 12761; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 289 BP; 37 A; 77 C; 102 G; 73 T; 0 other;
XX
XX Query Match 1.7%; Score 69.8; DB 22; Length 289;
XX Best Local Similarity 52.6%; Pred. No. 1.1e-07;
XX Matches 152; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
XX
XX 1758 CAGACTGTGAGACTGGGTGGGCGGAGATGATGCTTTGTCCCATGGCTGCGCCT 1817
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 289 CAGGACCCAAACAGCTTCAGACACTGACATGAAACATCAACCAACCCCT 230
XX
XX 1818 GGAAGCCCTGAGCAGCTGAGAGCTGACGACCGGAGAGTGTGAGATGAGG 1877
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 229 GAGAGCCCAAGACACGGGTGTCAGCCAACTGTGAGACACGAGTACTCTGGAGC 170
XX
XX 1878 AGCAGGCGGCTTCTCTGTGGAAAGCGCACCCCTCGTACTCTGTGACCAAC 1937

```

Db 169 GCCAGGCGGCTTCCTGCTTCAAAAAGAGAGCTGCTTAAGTACCTGCTCTTTACCGACGCCCC 110
Qy 1938 ACAGCACAAGTGGGGTGCACCCCGGGGCTACCGCATCGAGATGCTGAGTTTCTGGAG 1997
Db 109 AGAGAAACCCCTGGGGGACACAGCGCACGTACCGCTCGAGATCCATCTCATGGCGCACC 50
Qy 1998 AGCCGCTGCCCCAAMACAGCTCCATGCGCAGAGGCTTCAGCTGGAGAG 2046
Db 49 AGGTGCTGCCCCCGAGCTGCAGAGAGGAGCAGCATCACTGCGGCAAG 1

RESULT 15
ABA31594/c
ID ABA31594 standard; DNA: 289 BP.
XX ABA31594;
AC
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #10060 for gene expression analysis in human heart cell sample.
XX
KM Human: gene expression; heart; microarray; vascular system; probe;
KM cardiovascular disease; hypertension; cardiac arrhythmia;
KM congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID No 10060; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 289 BP; 37 A; 77 C; 102 G; 73 T; 0 other;

Query Match 1.7%; Score 69.8; DB 22; Length 289;
Best Local Similarity 52.6%; Pred. NO. 1.1e-07;
Matches 152; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 1758 CAGGACTGAGAACTGGGTGGGCGGAGATATGCTTTGCTCCCATGGCTGTGCCCT 1817

Db 289 CAGGACTGAGAACTGGGTGGGCGGAGATATGCTTTGCTCCCATGGCTGTGCCCT 230
Qy 1818 GAGCCCTGAGACACACAGCTGCAGAGCTGCAGTGCACCGAAGCTGCTGAGATGGAG 1877
Db 229 GAGGCCAAGACACCGGCTGTGTCAGCCAACTGTGAGCAGAGCAGTACTCTGCGGAGC 170
Qy 1878 AGCAGCCGCTTCTCTGTGGGAAGCCGACCCCTGCTTACTGTACCTGCGCAGCAAC 1937
Db 169 GCCAGGCGGCTTCCTGCTTCAAAAAGAGAGCTGCTTAAGTACCTGCTCTTTACCGACGCCCC 110
Qy 1938 ACAGCACAAGTGGGGTGCACCCCGGGGCTACCGCATCGAGATGCTGAGTTTCTGGAG 1997
Db 109 AGAGAAACCCCTGGGGGACACAGCGCACGTACCGCTCGAGATCCATCTCATGGCGCACC 50
Qy 1998 AGCCGCTGCCCCAAMACAGCTCCATGCGCAGAGGCTTCAGCTGGAGAG 2046
Db 49 AGGTGCTGCCCCCGAGCTGCAGAGAGGAGCAGCATCACTGCGGCAAG 1

Search completed: May 20, 2003, 03:47:38
Job time : 841 secs

Tue May 20 13:18:43 2003

us-10-081-408-1.rml

Page 1

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 03:31:13 ; Search time 214 Seconds
(without alignments)
5789.599 Million cell updates/sec

Title: US-10-081-408-1

Perfect score: 4040
Sequence: 1 gtcctccacccttagctc.....ccccaaaaaaaaaaaaa 4040

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 862724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptodata/2/lna/5a_COMB.seq:*
 - 2: /cgn2_6/ptodata/2/lna/5a_COMB.seq:*
 - 3: /cgn2_6/ptodata/2/lna/5a_COMB.seq:*
 - 4: /cgn2_6/ptodata/2/lna/5a_COMB.seq:*
 - 5: /cgn2_6/ptodata/2/lna/5a_COMB.seq:*
 - 6: /cgn2_6/ptodata/2/lna/5a_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.8	2.1	7218	1	US-08-232-463-14
2	51.6	1.3	4386	4	US-09-300-0088-1
3	49	1.2	289	4	US-09-007-005-17
4	49	1.2	289	4	US-09-244-796-17
5	46.4	1.1	1571	4	US-09-276-531-49
6	45	1.1	2061	2	US-08-960-022-11
7	44.2	1.1	421	1	US-08-480-784-24
8	44.2	1.1	421	1	US-08-483-553-24
9	44.2	1.1	421	1	US-08-487-002-24
10	44.2	1.1	421	1	US-08-483-554B-24
11	44.2	1.1	421	1	US-08-488-011B-24
12	44.2	1.1	421	1	US-08-850-727-24
13	44.2	1.1	421	4	PCT-US95-10202-24
14	44.2	1.1	421	5	PCT-US95-10203-24
15	44.2	1.1	421	5	PCT-US95-10220-24
16	43.6	1.1	7218	1	US-08-232-463-14
17	43	1.1	7210	2	US-08-257-963B-10
18	43	1.1	7210	4	US-08-367-841A-10
19	43	1.1	7210	5	PCT-US95-07201-10
20	43	1.1	14581	4	US-08-520-373D-4
21	43	1.1	22481	4	US-08-367-841A-43
22	43	1.1	22481	5	PCT-US95-07201-43
23	43	1.1	22484	4	US-09-875-223-2
24	42.4	1.0	53526	3	US-08-658-136-2
25	42.4	1.0	53577	3	US-08-658-136-1
26	42.4	1.0	4403765	4	US-09-103-840A-2
27	42.4	1.0	4411529	4	US-09-103-840A-1

c 28	42	1.0	12001	1	US-08-458-568A-11	Sequence 11, Appl
c 29	41.6	1.0	90050	4	US-09-245-041-5	Sequence 5, Appl
c 30	41	1.0	51259	3	US-08-781-891-209	Sequence 209, Appl
c 31	40.8	1.0	87350	3	US-08-781-891-79	Sequence 79, Appl
c 32	40.8	1.0	87543	3	US-09-791-211-3	Sequence 3, Appl
c 33	40.6	1.0	32042	4	US-09-245-281-44	Sequence 44, Appl
c 34	40.4	1.0	43795	4	US-08-742-185-101	Sequence 101, App
c 35	40	1.0	112132	4	US-09-741-150-3	Sequence 3, Appl
c 36	39.8	1.0	789	4	US-09-149-476-288	Sequence 288, App
c 37	39.8	1.0	3381	4	US-09-009-119-1	Sequence 1, Appl
c 38	39.8	1.0	3381	4	US-09-371-507-1	Sequence 1, Appl
c 39	39.8	1.0	3383	5	PCT-US95-09098-1	Sequence 1, Appl
c 40	39.8	1.0	3404	1	US-08-265-429A-1	Sequence 1, Appl
c 41	39.8	1.0	3404	5	PCT-US95-09069-1	Sequence 1, Appl
c 42	39.6	1.0	1056	4	US-09-524-162-1	Sequence 1, Appl
c 43	39.6	1.0	1335	2	US-08-985-090-3	Sequence 3, Appl
c 44	39.6	1.0	1335	2	US-09-165-543-3	Sequence 3, Appl
c 45	39.6	1.0	1335	3	US-09-167-354-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pL-F1s
US-08-232-463-14
Query Match 2.1%; Score 85.8; DB 1; Length 7218;

```
Best Local Similarity 2.48; Pred. No. 1.8e-11;
Matches 9; Conservative 250; Mismatches 122; Indels 0; Gaps 0;

QY 2866 AGAGTCTTCTGTCACATTCCTCTCAGTCAGGCTCTTCTCTCTCTCTCTCTC 2925
    ||| ||| | : : : : : : : : : : : : : : : : : : : : : : :
Db 1053 AGGAGCTTCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1112

QY 2926 TCACCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 2965
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1113 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1172

QY 2986 GAATCCGAGGAGATATCCATATGCCAGCCCTGTACTCCCGCCAGCTCATTTTC 3045
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1173 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1232

QY 3046 AGTCAGTCCGCTCTCTCTCTCAGCCCTATGAGAGTCTCAAGTCAGCGAGCCCTATC 3105
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1233 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1292

QY 3106 AGAGTGGCCAACTCCCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3165
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1293 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1352

QY 3166 GCTCCATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3225
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1353 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1412

QY 3226 TCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3246
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1413 YYYYYYYYYYYYYYYYYYYY 1433

RESULT 2
US-09-300-008B-1/c
; Sequence 1, Application US/09300008B
; Patent No. 6458534
; GENERAL INFORMATION:
; APPLICANT: Concannon et al.
; TITLE OF INVENTION: A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
; FILE REFERENCE: 9924-0003-228
; CURRENT APPLICATION NUMBER: US/09/300,008B
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: US 60/083,269
; PRIOR FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4386
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26)...(2287)
US-09-300-008B-1

Query Match 1.38; Score 51.6; DB 4; Length 4386;
Best Local Similarity 54.98; Pred. No. 0.005;
Matches 191; Conservative 0; Mismatches 144; Indels 13; Gaps 4;

QY 3701 TACATACATAGTATGAGCTAAAAAATATGATGCTTATATATATATATATAT 3760
    ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2830 TACAGCATTAAGGAAAAAATATTAATAACATTTATGATGATTCCTATATTTT 2771

QY 3761 AATCTTCAGTGAAGGAGCAGCAGCAAGAAATGTATTA-GCGCATCTCCATTTGT 3819
    ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 2770 AATCTTAACTAATAAAGGAGGTAGAACACATGATATGATATGATATTTGGT 2711

QY 3820 TTCGAAAGGAGTGAATATAACCATTAATTCCTATGATGCC-----TATTCGA 3872
    ||| | | | | | | | | | | | | | | | | | | | | | | |
Db 2710 AGGAAATATTATAAATAGGAATTTGAATAGGACTCAAAATCCCTGAAAGATTCATA 2651

QY 3873 ATAAATGGTAACTGATTTACTTTGGAGGAGGAAACCACTAGTTGAGAGAGAGAG 3932
```

```
Db 2650 AGAATAGATGATCATGATTTTCTTGAGGGAGGACCAAGTACGACAAAG---TG 2594

QY 3933 GAAGGCTTTAAACCTTACACCCCTTTTGTACATTTGAATTTGAACCATGTGAT 3992
    ||| | | | | | | | | | | | | | | | | | | | | | | |
Db 2593 GAAGGTGACTTATGCTTTTACCTTACCTACGATTAATTAATTTGAAC--TGTGACTAT 2536

QY 3993 TACTATTTCAAAATTAACATTAATGCGCCCAAAAAAAAAAAAAAAAAA 4040
    ||| | | | | | | | | | | | | | | | | | | | | | |
Db 2535 TAACTATTATTAACAAACATTTAAAAATGAAAAAAGATGCA 2488

RESULT 3
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihc
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc-feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.28; Score 49; DB 4; Length 289;
Best Local Similarity 6.18; Pred. No. 0.0059;
Matches 13; Conservative 103; Mismatches 97; Indels 0; Gaps 0;

QY 2865 AAGATGCTTCTGTCACATTCCTCTCAGTCAGGCTCTTCTCTCTCTCTCTCT 2924
    | | : : : : : : : : : : : : | | : : : : : : : : : :
Db 260 AYCICGCAVAYAGAYAGTYTAACGACGACGACGACGACGACGACGACGACGAC 201

QY 2925 CTCACCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2984
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 SYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYN 141

QY 2985 CGAATCCGAGGAGATATCCATATGCCAGCCCTGTACTGCCGCCGCCCTGATTT 3044
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 SYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYN 81

QY 3045 CAGTCAAGTCTGCTCTCTCTCTCTCAGCCCTATGG 3077
    : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 SYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYNMYSYN 48

RESULT 4
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihc
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
```


[illegible]

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-480-784-24

Query Match 1.1%, Score 44.2; DB 1; Length 421;
Best Local Similarity 50.5%; Pred.No. 0.11;
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0

QY 2870 TGCCTTCCTCATCATTCCTCTCATCCAGAGTCCTTCTCTCTGCTCTCTCTAC 2929
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 117 TCCCTTCCCTTCCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 176
QY 2930 CTACTTCCTCTCCCTCCCTCCCTCTGTCCTCTCTCTCTCTCTCTCTCTCTCAATTTCTCCGAAT 2989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 177 GTCATCCCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 236
QY 2990 CCTGAGGGGATATCCCTATATGCCAGACCCCTGTACTCTCCCGACGCTCATGTTTCAGTC 3049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 237 CCTTCTTCTCTCTCTCTCTCTCTCTCTTGTGACAGAGCTGTGCTGTCACACAGGCTGAGATG 296
QY 3050 AAGTCCGCTCTCTCTCC 3067
    ||| ||| ||| ||| |||
DB 297 CAGTGCCTGATCTCGNC 314

RESULT 8
US-08-483-553-24
: Sequence 24, Application US/08483553
: Patent No. 5709999
: GENERAL INFORMATION:
: APPLICANT: SKOLNICK, Mark H.
: APPLICANT: GOLDGAR, David E.
: APPLICANT: MIKI, Yoshio
: APPLICANT: SWENSON, Jeff
: APPLICANT: KAMP, Alexander
: APPLICANT: HARSHMAN, Keith D.
: APPLICANT: SHATTUCK-ELIDENS, Donna M.
: APPLICANT: TAVELIGIAN, Sean V.
: APPLICANT: WISEMAN, Roger W.
: TITLE OF INVENTION: Futreal, P. Andrew
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
: STREET: 1201 New York Avenue, N.W., Suite 1000
: City: Washington
: STATE: DC

```


NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-488-011B-24

Query Match 1.1%; Score 44.2; DB 1; Length 421;
Best Local Similarity 50.5%; Pred. No. 0.11;
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 2870 TGGCTGTGCACATCCGCTCCATCCAGGCTCTTCCTCTGCTCTCTCTCTAC 2929
DB 117 TCGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 176
QY 2930 CTACTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 2989
DB 177 GTCTACCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 236
QY 2990 CCTGAGGGGATATCCCTATCTCCAGCCCTGTACTCCCGCAGCCTCATTTTCA 3049
DB 237 CCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 296
QY 3050 AAGTCCGCTCTCTCC 3067
DB 297 CAGTGGCGTATCTCGNC 314

RESULT 12
US-08-850-727-24
Sequence 24, Application US/08850727
Patent No. 6162897

GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/850,727
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/483,554
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-850-727-24

Query Match 1.1%; Score 44.2; DB 4; Length 421;
Best Local Similarity 50.5%; Pred. No. 0.11;
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 2870 TGGCTGTGCACATCCGCTCCATCCAGGCTCTTCCTCTGCTCTCTCTAC 2929
DB 117 TCGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 176
QY 2930 CTACTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 2989
DB 177 GTCTACCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 236
QY 2990 CCTGAGGGGATATCCCTATCTCCAGCCCTGTACTCCCGCAGCCTCATTTTCA 3049
DB 237 CCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 296
QY 3050 AAGTCCGCTCTCTCC 3067
DB 297 CAGTGGCGTATCTCGNC 314

RESULT 13
PCR-US95-10202-24

Sequence 24, Application PC/TUS9510202

GENERAL INFORMATION:

APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Sliard, Jacques
APPLICANT: Emi, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Durocher, Francine
TITLE OF INVENTION: In Vivo Mutations and Polymorphisms
TITLE OF INVENTION: In the 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000

RESULT 14

```

PCT-US95-10203-24
: Sequence 24, Application PC/TUS9510203
: GENERAL INFORMATION:
: APPLICANT: Skolnick, Mark H.
: APPLICANT: Goldgar, David E.
: APPLICANT: Miki, Yoshio
: APPLICANT: Swenson, Jeff
: APPLICANT: Kamp, Alexander
: APPLICANT: Harshman, Keith D.
: APPLICANT: Shattuck-Eidens, Donna M.
: APPLICANT: Tavtigian, Sean V.
: APPLICANT: Wiseman, Roger W.
: APPLICANT: Futreal, P. Andrew
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
: STREET: 1201 New York Avenue, N.W., Suite 1000
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10203
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/409,305
: FILING DATE: 24-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/348,824
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08-308,104
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/300,266
: FILING DATE: 02-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,221
: FILING DATE: 12-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ihnen, Jeffrey L.
: REGISTRATION NUMBER: 28,957
: REFERENCE/DOCKET NUMBER: 24884-109347
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-962-4810
: TELEFAX: 202-962-8300
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 421 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
:
: PCT-US95-10203-24
:
: Query Match 1.1%; Score 44.2; DB 5; Length 421;
: Best Local Similarity 50.5%; Pred. No. 0.11;
: Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: May 20, 2003, 06:43:09 ; Search time 816 Seconds

(Without alignments)
6537.583 Million cell updates/sec

Title: US-10-081-408-1

Perfect score: 4040
Sequence: 1 gtcctccacccttagtc.....cccaaaaaaaaaaaaaa 4040

Scoring table:
Gapop 10.0, Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4040	100.0	4040	US-10-081-408-1	Sequence 1, Appl 1
2	4039.6	100.0	4040	US-09-954-456-521	Sequence 521, App
3	2209.2	58.7	3006	US-10-081-408-19	Sequence 19, Appl 1
4	1017.8	25.2	2572	US-09-919-497-1	Sequence 1, Appl 1
5	714.4	17.7	1051	US-09-917-800A-1473	Sequence 1473, App
6	422.2	10.5	2441	US-09-962-832-126	Sequence 126, App
7	401.4	9.9	4459	US-09-918-995-30382	Sequence 30382, A
8	315	7.8	4444	US-09-960-352-6153	Sequence 6153, App
9	310.2	7.7	4111	US-09-960-352-2265	Sequence 2265, App
10	305.4	7.6	397	US-09-960-352-2312	Sequence 2312, App
11	276	6.8	364	US-09-960-352-3111	Sequence 3111, App
12	275	6.8	388	US-09-960-352-5603	Sequence 5603, App
13	247.2	6.1	413	US-09-960-352-2524	Sequence 2524, App
14	209.8	5.2	405	US-09-918-995-35960	Sequence 35960, A
15	183.6	4.5	254	US-09-960-352-3010	Sequence 3010, App
16	154.6	3.8	507	US-09-728-445-97	Sequence 97, Appl 1
17	98.2	2.4	471	US-09-918-995-6654	Sequence 6654, App
18	87.6	2.2	382	US-09-960-352-7315	Sequence 7315, App
19	82.4	2.0	274	US-09-960-352-2062	Sequence 2062, App

20	74.4	1.8	54945	9	US-09-967-669-10	Sequence 10, Appl 1
21	69.8	1.7	289	10	US-09-864-761-16914	Sequence 16914, A
22	66.2	1.6	467	10	US-09-864-761-76	Sequence 76, Appl 1
23	51.6	1.3	744	9	US-09-989-920-64	Sequence 64, Appl 1
24	50.6	1.3	1691139	9	US-10-067-514-1	Sequence 1, Appl 1
25	48.2	1.2	6146	9	US-09-764-891-6229	Sequence 6229, App
26	47.8	1.2	726	10	US-09-864-761-19707	Sequence 19707, A
27	47.8	1.2	1951	10	US-09-864-761-2926	Sequence 2926, App
28	47.4	1.2	32012	9	US-09-764-891-8552	Sequence 8552, App
29	47.2	1.2	53226	10	US-09-818-264-3	Sequence 3, Appl 1
30	46.8	1.2	99014	10	US-09-880-107-3428	Sequence 3428, App
31	46.6	1.2	167343	10	US-09-962-436-281	Sequence 281, App
32	46.6	1.2	167343	10	US-09-964-824A-273	Sequence 273, App
33	46	1.1	31885	9	US-09-764-891-7211	Sequence 7211, App
34	46	1.1	31885	9	US-10-074-095-775	Sequence 775, App
35	46	1.1	31885	10	US-09-764-860-775	Sequence 775, App
36	46	1.1	31885	10	US-09-764-860-7530	Sequence 2530, App
37	46	1.1	31885	10	US-09-764-877-2541	Sequence 2541, App
38	45.8	1.1	442	10	US-09-880-107-2657	Sequence 2657, App
39	45.8	1.1	594	9	US-10-123-135-10	Sequence 10, Appl 1
40	45.8	1.1	24132	9	US-10-074-095-661	Sequence 661, App
41	45.8	1.1	24132	10	US-09-764-860-661	Sequence 661, App
42	45.8	1.1	32191	9	US-10-091-504-1955	Sequence 1955, App
43	45.8	1.1	32191	9	US-09-764-869-1955	Sequence 1955, App
44	45.4	1.1	334	9	US-09-796-692-2876	Sequence 2876, App
45	45.4	1.1	334	9	US-10-040-862-2876	Sequence 2876, App

ALIGNMENTS

RESULT 1
US-10-081-408-1
; Sequence 1, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrahms, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081,408
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4040
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (161)...(2449)
US-10-081-408-1
Query Match 100.0%; Score 4040; DB 9; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4040; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GTCCTTCCACCTTGTGTCACGATCTGATCCGGGAGACTCAGCCAGACTCCGGGA 60
DB 1 GTCCTTCCACCTTGTGTCACGATCTGATCCGGGAGACTCAGCCAGACTCAGCTCCGGGA 60
OY 61 GCCCCCCACCCGCTCCAGGAGCCACAGAGCCCCCTTGTGCTGGCGGTGATATCATG 120
DB 61 GCCCCCCACCCGCTCCAGGAGCCACAGAGCCCCCTTGTGCTGGCGGTGATATCATG 120
OY 121 CTCCTCTTGGTGAATCAGCTGCTCTTGTGCGGAAATGACAGAAAGCAATCCT 180
DB 121 CTCCTCTTGGTGAATCAGCTGCTCTTGTGCGGAAATGACAGAAAGCAATCCT 180

QY	181	CGNCTCTATTCTGGCCGATATCAACATCTTTCGTTGGTTGTGCTCGCGAGGG	240
Db	181	CGTCTCTATTCTGGCCGCTATCAACATCTTTCGTTGTGTGCTCGCTGGGG	240
QY	241	CAGGGGTGAGAGTGGGGGTGAACCCAGCCAGCTTCCCATTCGCCCTCTGATCTCCAG	300
Db	241	CAGGGGTGAGAGTGGGGGTGAACCCAGCCAGCTTCCCATTCGCCCTCTGATCTCCAG	300
QY	301	TGCCCACTTTGAGACACACCCCTGGCCAGAGCCAGCTGTTTGCAGACCTGAGCCGAGAGA	360
Db	301	TGCCCACTTTGAGACACACCCCTGGCCAGAGCCAGCTGTTTGCAGACCTGAGCCGAGAGA	360
QY	361	CGTACGGCTGTATGCGGTTTCTGACCCAGGGGCTGGGGCCAGGGGCTGGTGAATGACG	420
Db	361	CGTACGGCTGTATGCGGTTTCTGACCCAGGGGCTGGGGCCAGGGGCTGGTGAATGAGC	420
QY	421	CCAGGCCCGGCCCTCGGACAACTGTCTTCTCAGTGGAGTTGACAGCTGCCCTCCAGAGC	480
Db	421	CCAGGCCCGGCCCTCGGACAACTGTCTTCTCAGTGGAGTTGACAGCTGCCCTCCAGAGC	480
QY	481	TGCAGCCCTGGCTCACTTGGACAGGGGGAGCCCCACACTGCCCGGGAGGCACTGGCAT	540
Db	481	TGCAGCCCTGGCTCACTTGGACAGGGGGAGCCCCACACTGCCCGGGAGGCACTGGCAT	540
QY	541	CGTCTTTTGGCAGGCAACCCAGCCCAACTGATGATGAGCTGGTGGGGCCACTGCC	600
Db	541	CGTCTTTTGGCAGGCAACCCAGCCCAACTGATGATGAGCTGGTGGGGCCACTGCC	600
QY	601	TCACCCCTCTACATGCGGGAGTGACTGTGAGAGGCTACGGAGGCCCTCGCCCTATCA	660
Db	601	TCACCCCTCTACATGCGGGAGTGACTGTGAGAGGCTACGGAGGCCCTCGCCCTATCA	660
QY	661	CCGAGGCCCGGTGCTGTCTCCAGAGTACCTGGACATAGACAGATGATCTTCAACAGAGA	720
Db	661	CCGAGGCCCGGTGCTGTCTCCAGAGTACCTGGACATAGACAGATGATCTTCAACAGAGA	720
QY	721	CGTCCGCCCAGGCTTGTGGGCTTCTCCACCACTGTGTCTTCTCAAGACACCGGGACGGA	780
Db	721	CGTCCGCCCAGGCTTGTGGGCTTCTCCACCACTGTGTCTTCTCAAGACACCGGGACGGA	780
QY	781	CCTGGTACATGACACAGGCTCCCGGTGTGCAATCAGAGGACCGGGCACCTGGTT	840
Db	781	CCTGGTACATGACACAGGCTCCCGGTGTGCAATCAGAGGACCGGGCACCTGGTT	840
QY	841	TGGCCTTACTACAACATCTGGGGCGCTGGGTTCTTCTGACACACGTGGCTTGGAGCT	900
Db	841	TGGCCTTACTACAACATCTGGGGCGCTGGGTTCTTCTGACACACGTGGCTTGGAGCT	900
QY	901	GCTAGTGAACACAAAGGCCCTTGAACCTCGCCGCTGAGATATCCAGAAAGGTCTTATCA	960
Db	901	GCTAGTGAACACAAAGGCCCTTGAACCTCGCCGCTGAGATATCCAGAAAGGTCTTATCA	960
QY	961	AGGCGGCTACTACAGACAGCCTGGCCAGCGTGGAGGCCAGTTTGAAGCCGCGCTGTAA	1020
Db	961	AGGCGGCTACTACAGACAGCCTGGCCAGCGTGGAGGCCAGTTTGAAGCCGCGCTGTAA	1020
QY	1021	TGTGGTGTGATCCACAGAACTGGACAAAGTGGGCTCTGTCCTGAAGTCCCTGTGCC	1080
Db	1021	TGTGGTGTGATCCACAGAACTGGACAAAGTGGGCTCTGTCCTGAAGTCCCTGTGCC	1080
QY	1081	CCCGGGTCACTCCCTCTTACAGTTCTATCCCCAAGGCCCCCGCTTCAGTGTCAAGG	1140
Db	1081	CCCGGGTCACTCCCTCTTACAGTTCTATCCCCAAGGCCCCCGCTTCAGTGTCAAGG	1140
QY	1141	AAGTGAAGTGGCCCTCACTGTGAGATTTCTCTTGGCTCGGAGCATTTCAGTGGGCC	1200
Db	1141	AAGTGAAGTGGCCCTCACTGTGAGATTTCTCTTGGCTCGGAGCATTTCAAGTGGCC	1200
QY	1201	AAGGATCTTTAGCTTGTGCTTCCAAAGGAGAAAGTATTTAGAGATTAAGGCTCCAGA	1260
Db	1201	AAGGATCTTTAGCTTGTGCTTCCAAAGGAGAAAGTATTTAGAGATTAAGGCTCCAGA	1260
QY	1261	GGCCTTGGCATTATGATGAGAAATTCGCCAGCAGCAATGACAGCCCGCTATGTGATGG	1320

Db	1281	GGCCTTGGCCATCTATGGTGGAAATTCCCCACAGACATGAGACGCCCTATGGATGG	1320
Qy	1321	AGGCTTTGGCATGGGCAAGTACACACGCCCTGACCCGTTGGGTGACTGCCCCACTT	1380
Db	1321	AGGCTTTGGCATGGGCAAGTACACACGCCCTGACCCGTTGGGTGACTGCCCCACTT	1380
Qy	1381	GGCCACTACGTGACTGGGCACTTCTTTTGGAGTCCAGGCCCCCAAGACATAGCTGA	1440
Db	1381	GGCCACTACGTGACTGGGCACTTCTTTTGGAGTCCAGGCCCCCAAGACATAGCTGA	1440
Qy	1441	TGCCTTTTGGTGTGTTTGAACAGAACAGAGGCGCCCCCGGGGGGACACACTAGAGTCT	1500
Db	1441	TGCCTTTTGGTGTGTTTGAACAGAACAGAGGCGCCCCCGGGGGGACACACTAGAGTCT	1500
Qy	1501	CTACTCGCACTACTTTTGGGGGTCTTCGGGAAACGGTGTGGTCTGACATCTATGTCAC	1560
Db	1501	CTACTCGCACTACTTTTGGGGGTCTTCGGGAAACGGTGTGGTCTGACATCTATGTCAC	1560
Qy	1561	CTTGCTCACTATGACTATGTGTGGGATACGGTCTTCCACCCCACTGGGGCCATAGAAAT	1620
Db	1561	CTTGCTCACTATGACTATGTGTGGGATACGGTCTTCCACCCCACTGGGGCCATAGAAAT	1620
Qy	1621	ACGATTCATATGCCAGGGGGTACATACGCTGGGATTCCTTTGGTGTACTGGGAAGTA	1680
Db	1621	ACGATTCATATGCCAGGGGGTACATACGCTGGGATTCCTTTGGTGTACTGGGAAGTA	1680
Qy	1681	CGGGACCAAGTGTCAAGACACACCCCTGGGCAAGGTCCACACCACAGCCCACTTCAA	1740
Db	1681	CGGGACCAAGTGTCAAGACACACCCCTGGGCAAGGTCCACACCACAGCCCACTTCAA	1740
Qy	1741	GGTGATCTGGATGTACAGGACTGGAGAACTGGTCTGGGCCGAGATATGGTCTTGT	1800
Db	1741	GGTGATCTGGATGTACAGGACTGGAGAACTGGTCTGGGCCGAGATATGGTCTTGT	1800
Qy	1801	CCCCATGGCTGTGCCCTGGAGGCCCTAGACACCAAGCTGCAAGGCGAGGTGACCCGGAA	1860
Db	1801	CCCCATGGCTGTGCCCTGGAGGCCCTAGACACCAAGCTGCAAGGCGAGGTGACCCGGAA	1860
Qy	1861	GCTGCTGAGATGAGAGAGAGAGCGCGCTTCTCTGTGGAAAGCGCACCCCTCGCTACT	1920
Db	1861	GCTGCTGAGATGAGAGAGAGAGCGCGCTTCTCTGTGGAAAGCGCACCCCTCGCTACT	1920
Qy	1921	GTAACCTGGCCAGCAACACACAAAGTGGGCTCACCCCGGGGCTACCGCATTCAGAT	1980
Db	1921	GTAACCTGGCCAGCAACACACAAAGTGGGCTCACCCCGGGGCTACCGCATTCAGAT	1980
Qy	1981	GCTCAGCTTCTGTGGAGAGCGCGGCCCAAAACACTCCATGGGGAGAGGCTTCAGGTG	2040
Db	1981	GCTCAGCTTCTGTGGAGAGCGCGGCCCAAAACACTCCATGGGGAGAGGCTTCAGGTG	2040
Qy	2041	GGAGAGGTACAGCTGGCTGTGACCCAGCGGAAAGAGAGAGGCCAGTAGAGCAGCGGT	2100
Db	2041	GGAGAGGTACAGCTGGCTGTGACCCAGCGGAAAGAGAGAGGCCAGTAGAGCAGCGGT	2100
Qy	2101	TTTCAATCAGAAATGACCCCTTGGGGCCCCCACTGTGATTTTCAGTACTTCATCACAATGA	2160
Db	2101	TTTCAATCAGAAATGACCCCTTGGGGCCCCCACTGTGATTTTCAGTACTTCATCACAATGA	2160
Qy	2161	GACCACTTGTGGAAAGATTTTGGTGGCGTGGGAGACAGTGGTTTCTGCAATATCCACA	2220
Db	2161	GACCACTTGTGGAAAGATTTTGGTGGCGTGGGAGACAGTGGTTTCTGCAATATCCACA	2220
Qy	2221	TGCAGAGACATTTCTTAACACAGTGAATGGGGAAAGCGGTGGGCTTTCTCTCCGAC	2280
Db	2221	TGCAGAGACATTTCTTAACACAGTGAATGGGGAAAGCGGTGGGCTTTCTCTCCGAC	2280
Qy	2281	CTATATCTTTTGGAGCAAGACCCCTCTTCTACTCTGGCGACTCCATCTACTCCGAGG	2340
Db	2281	CTATATCTTTTGGAGCAAGACCCCTCTTCTACTCTGGCGACTCCATCTACTCCGAGG	2340
Qy	2341	GGACCAAGATGCTGGGGCTTGGAGGTCAACCCCTAGCTTGGCTTGGCCAGGCTGTGC	2400
Db	1281	GGCCTTGGCCATCTATGGTGGAAATTCCCCACAGACATGAGACGCCCTATGGATGG	1320

Db 2341 GGACAGAGATGCTGGGGGCTGAGAGTCAACCCCTAGCTTGCCTCCCAAGGCTGCTGC 2400
Qy 2401 CTGTGCCCCCACTCCTGCTTCTCTCCAGGGGGCTTCTCTCAACATAGGCGTCT 2460
Db 2401 CTGTGCCCCCACTCCTGCTTCTCTCCAGGGGGCTTCTCTCAACATAGGCGTCT 2460
Qy 2461 GGGAGGGGCAATGAGGCAAGGGCTCAAGGGCTGAGGGATGAGGGAGGAGCTGG 2520
Db 2461 GGGAGGGGCAATGAGGCAAGGGCTCAAGGGCTGAGGGATGAGGGAGGAGCTGG 2520
Qy 2521 GCAGTGGGCGGAGAGCTGTCTCTCTCTGCGAGAGACTCTCTTCTTCTCACTAC 2580
Db 2521 GCAGTGGGCGGAGAGCTGTCTCTCTCTGCGAGAGACTCTCTTCTTCTCACTAC 2580
Qy 2581 CTTCCCTGCAATCGGCTCTGTGAGCAGAGAGCTCTGACCTGTGATGCTGACAGAGG 2640
Db 2581 CTTCCCTGCAATCGGCTCTGTGAGCAGAGAGCTCTGACCTGTGATGCTGACAGAGG 2640
Qy 2641 GACACTGAACCTGTGTGATGCGAGCTGTATGAGTTCATGCTCAAGAGAGGCGAGCTGG 2700
Db 2641 GACACTGAACCTGTGTGATGCGAGCTGTATGAGTTCATGCTCAAGAGAGGCGAGCTGG 2700
Qy 2701 CCCAGCTGAGAGCCGTGGGCGAGGGCTTCCCTAGATGTTCCCTTGTGTGCTGTGCT 2760
Db 2701 CCCAGCTGAGAGCCGTGGGCGAGGGCTTCCCTAGATGTTCCCTTGTGTGCTGTGCT 2760
Qy 2761 TTCCCGAATCTTTTATGAGGCACTTCCAGAGCTCTTAAAGGGGGCTATTCCTGAGAGC 2820
Db 2761 TTCCCGAATCTTTTATGAGGCACTTCCAGAGCTCTTAAAGGGGGCTATTCCTGAGAGC 2820
Qy 2821 CCAGAGTAAAGGTTGCAAGTCCGCAAGTCCATGAGTGAAGTGAAGGATGTTCTGCTC 2880
Db 2821 CCAGAGTAAAGGTTGCAAGTCCGCAAGTCCATGAGTGAAGTGAAGGATGTTCTGCTC 2880
Qy 2881 ACATTCCTCTCATCAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Db 2881 ACATTCCTCTCATCAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Qy 2941 TCT 3000
Db 2941 TCT 3000
Qy 3001 ATCCCTATGTCAGAGCCCTGTGACTCCGCCAGCCCTGAGTTTCAAGTCAATTCGCT 3060
Db 3001 ATCCCTATGTCAGAGCCCTGTGACTCCGCCAGCCCTGAGTTTCAAGTCAATTCGCT 3060
Qy 3061 CCTCTCAAGCCCTATGAGAGTCTCAAGGTCAAGGGAGCCCTATCAGATGAGTGGCAATCC 3120
Db 3061 CCTCTCAAGCCCTATGAGAGTCTCAAGGTCAAGGGAGCCCTATCAGATGAGTGGCAATCC 3120
Qy 3121 TGTGTGTGCTTCT 3180
Db 3121 TGTGTGTGCTTCT 3180
Qy 3181 CTGGGGCTGGGTGTCTTTCAGAGAGCTCTCTGTCATTTGTGTGCTGCTGCTGCTGCTG 3240
Db 3181 CTGGGGCTGGGTGTCTTTCAGAGAGCTCTCTGTCATTTGTGTGCTGCTGCTGCTGCTG 3240
Qy 3241 CTCTCTATAGAGAGAGATGCTATGCTGACAGAGAGCTCAAGTATGAGTTCAGTGA 3300
Db 3241 CTCTCTATAGAGAGAGATGCTATGCTGACAGAGAGCTCAAGTATGAGTTCAGTGA 3300
Qy 3301 TTTGGGGGTGCAATATATGAGAGATGGCCATTTTGTACAGGGGCTGTATTTCTGCA 3360
Db 3301 TTTGGGGGTGCAATATATGAGAGATGGCCATTTTGTACAGGGGCTGTATTTCTGCA 3360
Qy 3361 CAGCCTGTTTGGAGAGCTGAGTGAACCAAGGGGTGGGATCAAGATGAGAGAGCA 3420
Db 3361 CAGCCTGTTTGGAGAGCTGAGTGAACCAAGGGGTGGGATCAAGATGAGAGAGCA 3420
Qy 3421 GCCCCTACAACTCAGAGCCAGAGGAGGAGGCTGTCTCAATTCAGGAGTGCAG 3480
Db 3421 GCCCCTACAACTCAGAGCCAGAGGAGGAGGCTGTCTCAATTCAGGAGTGCAG 3480

Qy 3481 ATGAGCTGGCCCTGGGTGAGAGTGGGGTCTGCGCTAGTGGGGAGGGGCTGGCTGGGT 3540
Db 3481 ATGAGCTGGCCCTGGGTGAGAGTGGGGTCTGCGCTAGTGGGGAGGGGCTGGCTGGGT 3540
Qy 3541 GGGGAGGGGCTGGGCTGGTCCAGGCTTGGGCTGATTCATTCATCTGCTGCTGCTGCTG 3600
Db 3541 GGGGAGGGGCTGGGCTGGTCCAGGCTTGGGCTGATTCATTCATCTGCTGCTGCTGCTG 3600
Qy 3601 AGGTGATGATGGGAGGAGCAAGAAATGCAAGAGATGAGAGAGCAAGAACTTT 3660
Db 3601 AGGTGATGATGGGAGGAGCAAGAAATGCAAGAGATGAGAGAGCAAGAACTTT 3660
Qy 3661 TTTCTCTAAGAGCTGTGTTAATCAATTCATGATACAGCTTCAATTAATATGATGAG 3720
Db 3661 TTTCTCTAAGAGCTGTGTTAATCAATTCATGATACAGCTTCAATTAATATGATGAG 3720
Qy 3721 CTAAATAATATGATGCTTATATGATATGATATGATATGATATGATATGATATGAT 3780
Db 3721 CTAAATAATATGATGCTTATATGATATGATATGATATGATATGATATGATATGAT 3780
Qy 3781 AAGCCAGAAATGTGTATAGCGCACTTCCATTTGTGTGTGAGAGAGAGATATATA 3840
Db 3781 AAGCCAGAAATGTGTATAGCGCACTTCCATTTGTGTGTGAGAGAGAGATATATA 3840
Qy 3841 AACACATATGCTTATGATGCTTATGATGATGATGATGATGATGATGATGATGATG 3900
Db 3841 AACACATATGCTTATGATGCTTATGATGATGATGATGATGATGATGATGATGATG 3900
Qy 3901 GAGGGAGACGATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Db 3901 GAGGGAGACGATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Qy 3961 TACATTTTGAATTTGAACCATGATGATGATGATGATGATGATGATGATGATGATG 4020
Db 3961 TACATTTTGAATTTGAACCATGATGATGATGATGATGATGATGATGATGATGATG 4020
Qy 4021 CCC 4040
Db 4021 CCC 4040

RESULT 2
US-09-954-456-521
Sequence 521, Application US/09954456
Patent No. US20020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
CURRENT FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27

D	901	GCTAGTAAACACAGAGGCCCTT	9ACCCTCCCGCTGAGCACTACGAAGGTCTCTATCA	960
Q	961	AGGCGGCTACTACGACAGCCTG	9GCCAGCTGGAGGCCCAAGTTTGAGGCGGCGCTGTGAA	1020
D	961	AGGCGGCTACTACGACAGCCTG	9GCCAGCTGGAGGCCCAAGTTTGAGGCGGCGCTGTGAA	1020
Q	1021	TGTGTGCTCTATCCACAGAAAT	9GGACAGAGTGGGTGTCGTGTGCTCCCTGAAGTCCCTGTGCC	1080
D	1021	TGTGTGCTCTATCCACAGAAAT	9GGACAGAGTGGGTGTCGTGTGCTCCCTGAAGTCCCTGTGCC	1080
Q	1081	CCCGGGTCCAGCTCCCGCTCT	9ATACAGTTCTATCCCAAGGCCCGCCCTCTACGTGCCAGG	1140
D	1081	CCCGGGTCCAGCTCCCGCTCT	9ATACAGTTCTATCCCAAGGCCCGCCCTCTACGTGCCAGG	1140
Q	1141	AAGTCAGTGGCCTCCTC	9ACTGTGAGACTTTCTCTTTGGCCTCGSAGCATTCAGTGGCC	1200
D	1141	AAGTCAGTGGCCTCCTC	9ACTGTGAGACTTTCTCTTTGGCCTCGSAGCATTCAGTGGCC	1200
Q	1201	AAGATCTTTAGCGTTGCGCTT	9CCAAAGAGAAAGACTAGTTTATGAGTAAACCTTC	1260
D	1201	AAGATCTTTAGCGTTGCGCTT	9CCAAAGAGAAAGACTAGTTTATGAGTAAACCTTC	1260
Q	1261	GGCCTTGCCCATCTATAGT	9GTGAATTCGCCAGAGCAATGAGACCCCTATGTGGATGG	1320
D	1261	GGCCTTGCCCATCTATAGT	9GTGAATTCGCCAGAGCAATGAGACCCCTATGTGGATGG	1320
Q	1321	AGGCTTTGGCATGGGCAAG	9ATCACAGGCCCGTGAGCCGCTGGGGTGGAGCTGCCCTACTT	1380
D	1321	AGGCTTTGGCATGGGCAAG	9ATCACAGGCCCGTGAGCCGCTGGGGTGGAGCTGCCCTACTT	1380
Q	1381	GGCCACCTAGCTGGAC	9TGCCCTTTCTTTGGAGTCCAGGCGCCCAAGACATACGTGA	1440
D	1381	GGCCACCTAGCTGGAC	9TGCCCTTTCTTTGGAGTCCAGGCGCCCAAGACATACGTGA	1440
Q	1441	TGCGCTTTGTGTGTTTGA	9AACAGAACAGGGCGTCCCGCGGGGGACACACACTAGACT	1500
D	1441	TGCGCTTTGTGTGTTTGA	9AACAGAACAGGGCGTCCCGCGGGGGACACACACTAGACT	1500
Q	1501	CTACTCGCACTACTTTGG	9GGGTCTTGCGGAAACGAGTGTGTCAGATCTATGTCCAC	1560
D	1501	CTACTCGCACTACTTTGG	9GGGTCTTGCGGAAACGAGTGTGTCAGATCTATGTCCAC	1560
Q	1561	CTTGCTCAACTATATGTG	9GGGATACGCTGCTTCACCCCACTGGGGCCATAGAAAT	1620
D	1561	CTTGCTCAACTATATGTG	9GGGATACGCTGCTTCACCCCACTGGGGCCATAGAAAT	1620
Q	1621	ACGATTCATATCCACGG	9GCTACAGCTCGGCAATTCCTTTGGTGTACTAGCGGAATA	1680
D	1621	ACGATTCATATCCACGG	9GCTACAGCTCGGCAATTCAGCTGCGATTCCTTTGGGTGCTAC	1680
Q	1681	CGGGAACCAAGTGTCA	9GACACACCCACAGCGCCCACTTCAA	1740
D	1681	CGGGAACCAAGTGTCA	9GACACACCCACAGCGCCCACTTCAA	1740
Q	1741	GGTGGATGTGATATAC	9AGACAGCTGGGAACTGGGTCTGGCCGAGATATGTCTTGT	1800
D	1741	GGTGGATGTGATATAC	9AGACAGCTGGGAACTGGGTCTGGCCGAGATATGTCTTGT	1800
Q	1801	CCCCATGGCTGTGCC	9CTGGAGCCCTGAGCACAGCTGCAGAGGCTGCAAGTACCCGGA	1860
D	1801	CCCCATGGCTGTGCC	9CTGGAGCCCTGAGCACAGCTGCAGAGGCTGCAAGTACCCGGA	1860
Q	1861	GCTGCTGGAATGAG	9AGAGAGGCGGCTTCTCTGTGGGAAACGCCACCCCTCGTACT	1920
D	1861	GCTGCTGGAATGAG	9AGAGAGGCGGCTTCTCTGTGGGAAACGCCACCCCTCGTACT	1920
Q	1921	GTACCTGGCCAGCAAC	9ACAGAAAGTGTGGGTCAACCCCGGGGCTTACCGCATCCAGAT	1980
D	1921	GTACCTGGCCAGCAAC	9ACAGAAAGTGTGGGTCAACCCCGGGGCTTACCGCATCCAGAT	1980
Q	1981	GCTCAGCTTTGCTG	9AAGCGGCTGCCCAAAACAGCTCCATGTGGAGAGGCTTACGTG	2040
D	1981	GCTCAGCTTTGCTG	9AAGCGGCTGCCCAAAACAGCTCCATGTGGAGAGGCTTACGTG	2040


```

: TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
: FILE REFERENCE: 13425-03501
: CURRENT APPLICATION NUMBER: US/10/081,408
: CURRENT FILING DATE: 2002-02-21
: PRIOR APPLICATION NUMBER: SE 0100625-3
: PRIOR FILING DATE: 2001-02-23
: PRIOR APPLICATION NUMBER: US 60/272,247
: PRIOR FILING DATE: 2001-02-28
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 19
: LENGTH: 3006
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Recombinant construct
: NAME/KEY: CDS
: LOCATION: (1)...(2994)
: US-10-081-408-19

Query Match      54.7% Score 2209.2; DB 9; Length 3006;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2211; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 245 GGTGAGATGGGGTGAACCCAGCAGCTTCCCATTCCTCTGTATCTCCAGTCC 304
DB 790 GGTGAGATGGGGTGAACCCAGCAGCTTCCCATTCCTCTGTATCTCCAGTCC 849
QY 305 CAGCTTGGACACACCTGGCCAGAGCAGCTGTTTGACAGCTAGAGCCGAGAGAGCTG 364
DB 850 CAGCTTGGACACACCTGGCCAGAGCAGCTGTTTGACAGCTAGAGCCGAGAGAGCTG 909
QY 365 ACGGCTGTGATGCGCTTCTGACCCAGGCGTGGGGCCAGGGGCTGTGATGACAGCCAG 424
DB 910 ACGGCTGTGATGCGCTTCTGACCCAGGCGTGGGGCCAGGGGCTGTGATGACAGCCAG 969
QY 425 GCCCGGCCCTGGACAACTGTCTTCTCAGTGTGAGTGTGAGCTCTCCCAAGGCTGCA 484
DB 970 GCCCGGCCCTGGACAACTGTCTTCTCAGTGTGAGTGTGAGCTCTCCCAAGGCTGCA 1029
QY 485 GCCCTGCTCACTTGTGAGAGGGGAGCCGCCACCTGCCCGGAGGACACTGGCCATGCTC 544
DB 1030 GCCCTGCTCACTTGTGAGAGGGGAGCCGCCACCTGCCCGGAGGACACTGGCCATGCTC 1089
QY 545 TTCTTGGCAGGCAACCCAGCCCAAGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 604
DB 1090 TTCTTGGCAGGCAACCCAGCCCAAGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1149
QY 605 CCTCTCACTGCGGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 664
DB 1150 CCTCTCACTGCGGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1209
QY 665 CGCCCGGCTGTTCACAGAGTACCTGAGACATAGACAGATGATTTCAACAGAGAGCTG 724
DB 1210 CGCCCGGCTGTTCACAGAGTACCTGAGACATAGACAGATGATTTCAACAGAGAGCTG 1269
QY 725 CCCCAGGCTTCTGGGCTTCTGACACACTGTGCTTCTACAAAGCAGGGGAGGAGAACTG 784
DB 1270 CCCCAGGCTTCTGGGCTTCTGACACACTGTGCTTCTACAAAGCAGGGGAGGAGAACTG 1329
QY 785 GTGACAATGACACAGGCTGCCCGTGTGCTGCAATCAGGGGAGCCGGGCAACCTGGTTTGGC 844
DB 1330 GTGACAATGACACAGGCTGCCCGTGTGCTGCAATCAGGGGAGCCGGGCAACCTGGTTTGGC 1389
QY 845 CTCTACTACAACTCTGGGAGCTGGGCTTCTCTGACACAGCTGGGCTGTGAGCTGCTA 904
DB 1390 CTCTACTACAACTCTGGGAGCTGGGCTTCTCTGACACAGCTGGGCTGTGAGCTGCTA 1449
QY 905 GTGAACCAAGAGCCCTTGACCTGCGCTGGAGCTATTCAGAGAGGCTTCTATCAAGGC 964
DB 1450 GTGAACCAAGAGCCCTTGACCTGCGCTGGAGCTATTCAGAGAGGCTTCTATCAAGGC 1509
QY 965 CGCTACTAGCAGAGCTGGCCAGCTGAGAGGCCAGTTTGAAGCCGGCTGTGATGTG 1024
```

```

DB 1510 CGCTACTAGCAGAGCTGGCCAGCTGGAGGCCAGATTTGAGGGCGCTGTGTAATGTG 1569
QY 1025 GTGCTGATCCAGACAAATGAGCAGAGTGGAGTCTGTGCTCCGGAAGTCCCTGTGCCCCG 1084
DB 1570 GTGCTGATCCAGACAAATGAGCAGAGTGGAGTCTGTGCTCCGGAAGTCCCTGTGCCCCG 1629
QY 1085 GGTCCAGCTCCCTCTACAGTCTATCCCAAGGCCCGCTTCAAGTGTCCAGGAGT 1144
DB 1630 GGTCCAGCTCCCTCTACAGTCTATCCCAAGGCCCGCTTCAAGTGTCCAGGAGT 1689
QY 1145 CGAGTGGCTCTCTCACTGTGACTTCTCTTGTGGCTCGGAGCATTCAGTGGCCCAAG 1204
DB 1690 CGAGTGGCTCTCTCACTGTGACTTCTCTTGTGGCTCGGAGCATTCAGTGGCCCAAG 1749
QY 1205 ATCTTGAAGTCTGCTTCCAGAGAGAAAGACTAGTTATGATTAAGCTCCAGAGGCC 1264
DB 1750 ATCTTGAAGTCTGCTTCCAGAGAGAAAGACTAGTTATGATTAAGCTCCAGAGGCC 1809
QY 1265 TTGGCATCTATGTGTGAATTTCCCAAGCAGCAATGACGACCCGCTATGTGATGAGGC 1324
DB 1810 TTGGCATCTATGTGTGAATTTCCCAAGCAGCAATGACGACCCGCTATGTGATGAGGC 1869
QY 1325 TTTGGCATGGGCAAGTACACAGCCCTGACCCGCTGGAGTGTGACTGCCCTACTTGGCC 1384
DB 1870 TTTGGCATGGGCAAGTACACAGCCCTGACCCGCTGGAGTGTGACTGCCCTACTTGGCC 1929
QY 1385 ACCTAAGTGGAGCTGGCACTTCTTTTGGAGTCCAGAGCCGCCCAAGACAAATACGATGCC 1444
DB 1930 ACCTAAGTGGAGCTGGCACTTCTTTTGGAGTCCAGAGCCGCCCAAGACAAATACGATGCC 1989
QY 1445 TTTTGTGTGTTGAACAGAACAGAGGCTCCCTCGCGGACACACACTGATCTCTAC 1504
DB 1990 TTTTGTGTGTTGAACAGAACAGAGGCTCCCTCGCGGACACACACTGATCTCTAC 2049
QY 1505 TCGCATCTATTGGGGGCTTGGGAAAGGAGTGCAGTGCAGATCTATCCACCTG 1564
DB 2050 TCGCATCTATTGGGGGCTTGGGAAAGGAGTGCAGTGCAGATCTATCCACCTG 2109
QY 1565 CTCAACTATGACTATGTGTGAGTACGGTCTTCCACCCCAAGTGGGGCCATGAAATACGA 1624
DB 2110 CTCAACTATGACTATGTGTGAGTACGGTCTTCCACCCCAAGTGGGGCCATGAAATACGA 2169
QY 1625 TTCTATGCGCAGGGCTATATAGCTGGGCAATTCCTTGTGCTAGTCTGGAATACAGG 1684
DB 2170 TTCTATGCGCAGGGCTATATAGCTGGGCAATTCCTTGTGCTAGTCTGGAATACAGG 2229
QY 1685 AACCAAGTGTGAGAGCAGCAGCTGGGACAGGTTCCACACCCACAGGCCCCACTTCAAGTG 1744
DB 2230 AACCAAGTGTGAGAGCAGCAGCTGGGACAGGTTCCACACCCACAGGCCCCACTTCAAGTG 2289
QY 1745 GATCTGATGTAGCAGAGCTGGAGAACTGGGTCTGGGCCGAGATATGTCTTGTCCCC 1804
DB 2290 GATCTGATGTAGCAGAGCTGGAGAACTGGGTCTGGGCCGAGATATGTCTTGTCCCC 2349
QY 1805 ATGCTGTGCCCTGGAGCCCTGAGCAGCAGCTGCAAGAGCTGCAAGTACCCGGAAGCTG 1864
DB 2350 ATGCTGTGCCCTGGAGCCCTGAGCAGCAGCTGCAAGAGCTGCAAGTACCCGGAAGCTG 2409
QY 1865 CTGAGATGTGAGAGCAGAGCGGCTCTCTGTTGGGAAAGCGCACCCCTCGCTACTGCTGAC 1924
DB 2410 CTGAGATGTGAGAGCAGAGCGGCTCTCTGTTGGGAAAGCGCACCCCTCGCTACTGCTGAC 2469
QY 1925 CTGGCAGACCAACAGCAGCAAGTGGGCTACCCCGGGGCTACCGCATCCAGATGCTC 1984
DB 2470 CTGGCAGACCAACAGCAGCAAGTGGGCTACCCCGGGGCTACCGCATCCAGATGCTC 2529
QY 1985 AGCTTGTCTGAGAGCCGCTGCCAAAACAGCTCATGCGAGAGGCTTCACTGGGAG 2044
DB 2530 AGCTTGTCTGAGAGCCGCTGCCAAAACAGCTCATGCGAGAGGCTTCACTGGGAG 2589
QY 2045 AGGTACAGCTGGCTGTACCCAGAGGAGGAGGAGGCCAGTACAGCAGGCTTTC 2104
```

Db 2590 AGTACACGCTGGCTGTGTGACCCAGGGAAGAGAGAGACCCAGTAGACAGACGCTTTTC 2649
Qy 2105 AATCAGATGACCCCTTGGGCCCCCACTGTGATTTTCAGTACTTATCAACATGAGACC 2164
Db 2650 AATCAGATGACCCCTTGGGCCCCCACTGTGATTTTCAGTACTTATCAACATGAGACC 2709
Qy 2165 ATTTGTGGAAGGATTTGGTGGGCTGGGAGACAGCTGGTTTCTGATATCCACATGGA 2224
Db 2710 ATTTGTGGAAGGATTTGGTGGGCTGGGAGACAGCTGGTTTCTGATATCCACATGGA 2769
Qy 2225 GAGGACATCTCTAACACAGTACTGTGGGGAACGGGCTGGGCTTCTCTCCGACCTAT 2284
Db 2770 GAGGACATCTCTAACACAGTACTGTGGGGAACGGGCTGGGCTTCTCTCCGACCTAT 2829
Qy 2285 AACTCTTTGACGAGAGACCCCTCTCTACTCTGCGGACTCCATCTACTCTCCAGGGGAC 2344
Db 2830 AACTCTTTGACGAGAGACCCCTCTCTACTCTGCGGACTCCATCTACTCTCCAGGGGAC 2889
Qy 2345 CAGGATGCTGGGGCTGGGAGGTCAACCCCTAGCTTGGCTGGCCAGGCTGGCTGT 2404
Db 2890 CAGGATGCTGGGGCTGGGAGGTCAACCCCTAGCTTGGCTGGCCAGGCTGGCTGT 2949
Qy 2405 GCGCCGACCTCCCTCTCTCTCCACGAGGGGCTTCTCTACAACTAGGCGGTG 2458
Db 2950 GCGCCGACCTCCCTCTCTCTCCACGAGGGGCTTCTCTACAACTAGTGAAGTC 3003

RESULT 4

US-09-919-497-1
; Sequence 1, Application US/09919497
; Patent No. US2002010662A1
; GENERAL INFORMATION:
; APPLICANT: Muller, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 1
; LENGTH: 2572
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-497-1

Query Match 25.2%; Score 1017.8; DB 10; Length 2572;
Best Local Similarity 67.6%; Pred. No. 8.5e-294;
Matches 1545; Conservative 0; Mismatches 637; Indels 105; Gaps 4;

Qy 161 ATACACAGAGACAACTCTCTGCTCTCTATCTGCGCTCATACCATCTTTGCTTG 220
Db 27 ATGATCTCTCAAGTATGCTCTGCGCTCTGCTGCTCTCTCTCTCTCTCTCTCTCT 86
Qy 221 GTTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280
Db 87 GCTTATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 134
Qy 281 TGGCCCTCTCTATCTCTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 340
Db 135 TGGCCCTCTCTATCTCTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 194
Qy 341 GCAGACCTGAGCCGAGAGAGAGCTGAGGCTGTATGCGCTTCTTGTGACCCAGCGGCTGGG 400
Db 195 GCAGACCTGAGCCGAGAGAGAGCTGAGGCTGTATGCGCTTCTTGTGACCCAGCGGCTGGG 254
Qy 401 CGAGGCTGTGATGAGCCGAGGCGGCTCTGAGACAACTGTGCTCTCTCTCTCTCTCTCT 460
Db 255 CGAGGCTGTGATGAGCCGAGGCGGCTCTGAGACAACTGTGCTCTCTCTCTCTCTCTCT 314
Qy 461 TTGACGCTGCTCTCAAGGCTGAGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 520
Db 1389 GTGCTGAGGTCTGTGCTATCTGTGGCACTATGACTATCTGTGGGACTTGTGTGTAC 1448

Db 315 CTGCAAGCTGCCCCCAAGGCTGAGCCCTTGCCCACTGTGACAGGGGAGCCCCCACT 374
Qy 521 GCCCGGAGGCACTGGCCATCTCTTTTGGCAGGCAACCCCACTGAGTGTG 580
Db 375 GCCCGGAGGCACTGGCCATCTCTTTTGGCAGGCAACCCCACTGAGTGTG 434
Qy 581 CTGCTGTGGGCACTGGCCATCTCTTTTGGCAGGCAACCCCACTGAGTGTG 640
Db 435 CTGCTGTGGGCACTGGCCATCTCTTTTGGCAGGCAACCCCACTGAGTGTG 494
Qy 641 GGAGGCCCCCTGCTCTATACCGACGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 700
Db 495 GGAGGCCCCCTGCTCTATACCGACGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 554
Qy 701 CAGATGATTTCAACAGAGAGCTGCCCCAGGCTTCTGCGCTTCTCTCTCTCTCTCTCT 760
Db 555 AGGCATCTCAAGATGTGAGAGTACCCCAAGGCTCTCTCTCTCTCTCTCTCTCTCTCT 608
Qy 761 TACAAACCGGGGAGGAGAACCTGTGACAAATGACACGAGGCTCTCTCTCTCTCTCTCT 820
Db 609 TTCACTTCAAAATGCTCTTACCTGCGAGCTGTGACACCCCTGCGGCTTGGCTG 668
Qy 821 GGGGACCGGGCACTGCTGCTTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 880
Db 669 AGGGAAGCAACTACCTGAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 728
Qy 881 CACCACTGTGGCTTGTGAGCTGCTGAGAACACAGAGCCCTTGACCTTGGCTGCTG 940
Db 729 CACCCGCTGGGCTGTGAGCTACTGACACAGGCGCTGAGCCCTGCTGCTGCTGCTG 788
Qy 941 ATCCAAAGTGTCTTATCAAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1000
Db 789 GTCCACAGGCTTCTTACCTTGGGCACTATGACAGCTGTGGGCTGAGGAGCGGAG 848
Qy 1001 TTGAGAGCGGCTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1060
Db 849 TTGAGAGCGGCTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 908
Qy 1061 TCCCTGAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1120
Db 909 TCCCTGAGTCTGGAAGCTCTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 968
Qy 1121 CCGGCTGAGTCTGAGGAGGCTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1180
Db 969 TCCCTGAGTCTGAGGAGGCTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1028
Qy 1181 CTGAGAGCTTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1240
Db 1029 CATGGGCTGTGAGCGGCTGAGGATTTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1088
Qy 1241 TATGAGTATGCTCTCAAGAGGCTTGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1300
Db 1089 TATGAGTATGCTCTCAAGAGGCTTGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1148
Qy 1301 ACAGCCGCTATGCTGAGGAGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1360
Db 1149 CTGACTGCTATTTGATGATGACACTTTGAGCTGCGGCTGATCAACGCGAGCTTGTG 1208
Qy 1361 GGGGTGAGTCCCTCTACTTGGCCACCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1420
Db 1209 GAGTGTGAGTCCCTCTCAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1268
Qy 1421 GCGCCCAAGCAATAGCTGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1480
Db 1269 GCACTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1328
Qy 1481 CGGCGACACCTCAAGTCT 1540
Db 1329 CGAAGGCAACCAATTAATCTCTCAAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1388
Qy 1541 GTGCTGAGTCTATGCTCAAGCTTGTCTCAATATGATATGCTGCTGCTGCTGCTGCTGCT 1600
Db 1389 GTGCTGAGGTCTGTGCTATCTGTGGCACTATGACTATCTGTGGGACTTGTGTGTAC 1448


```
QY 1506 CGCACTACTTGGGGGCTCTTGGGAAACGGTGTGCTGATATCTATGTCACCTTGC 1565
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1414 TCACTTCTATGCGGGCTGAAAGGGCCAGAGTCTGTGTGGGACACTTAACTGTCT 1473
QY 1566 TCACTATGACTATGTGTGGGATACGGTCTTCCACCCAGTGGGGCCATAGAAATACGAT 1625
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1474 ACAATTATGATATGTTGGGACTTATCTTACTTACCCCAAGGGGGATGAGGGCCACA 1533
QY 1626 TCTATGCCACGGGCTACATCACTGCGCATTCCTTTGTGCTCTACTGGGAAGTACGGGA 1685
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1534 TGCATGCCACGTGGCTACCTCCACGGCCACTTCTACACCCCGAGGGGGCTGCCCGCGCA 1593
QY 1686 ACCAAGTGTCAAGACGACACCCCTGGGACGGTCCACACCCAGCCGCGCTCAAGGTGG 1745
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1594 CTCGCCCTGCACACCACTGATTTGGCAACATACACTACTTGTGCTACCTACCGGTAG 1653
QY 1746 ATCTGATATGACAGCACTGGAATGAGTCTGGCCGAGATATGCTTCTTGTCCCA 1805
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1654 ACCTGATGTGGACGACCAAGAACAGCTTCCAGACACTGCAATGAAAGCTAGAAAAA 1713
QY 1806 TGGCTGTGCTGGAGCCCTGAGCACAGCTGACAGAGCTGCAAGTGCACCCGAAAGCTGC 1865
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1714 TCACCAACCCCTGGAGCCCAAGACACGCGTGTCCAGCAACTGTGGAGCAGCGCACT 1773
QY 1866 TGGAGATGGAGAGCAGCCGCTTCTGTGGGAAGCGCACCCCTGCTACTGCTTAC 1925
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1774 ACTCCTGGAGAGGCGGAGCGGCTTCCGCTTAAAGAAAGTGCCTAAGTACCTGTCT 1833
QY 1936 TGGCAGCAACCAAGCAAGAGTGGGTGACCCCGGGGCTACCGCATCCAGATGCTCA 1985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1834 TTACAGCCCGCAGAGAGAACCCCTGGGGGACACAGCGAGCTAACGGCTGACAGATCACT 1893
QY 1986 GCTTGTGTGAGAGCCGCTTCCCAAAACAGTCCATGGCGAGAGCTTCACTGGGAGA 2045
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1894 CCATGGCCGACAGGTGTGCTGCCCCAGGCTGGCAGAGAGAGGACCACTCTGGGCAA 1953
QY 2046 GSTACCACTGGCTGTGACCCAGGAAAGAGAGAGAGCCAGTAGACAGCAGCTTTTCA 2105
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1994 GSTACCCCTGGCAGTACAGCAAGTACCGGAGTCCGAGCTGTGCAGCAGCAGATCTACC 2013
QY 2106 ATCAGATGACCTTGGGCCCCCACTGTGATTTGATGACTTCAAT--CAACAATGAGA 2162
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2014 ACCAAGACGACCCCTGGACCCGCGCTGTGTAGAGATTTCTTCAACAACAGAGA 2073
QY 2163 CCATTGTGGAAGGATTTGGTGGCTGGGTGACAGCTGTCTTGTGCAATCCACAGAT 2222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2074 ACATGAAATGAGGACCTGGTGGCTGTGGGTAGCGTGGCTTCTGACATCCCCACT 2133
QY 2223 CAGAGACATTCCTTAACACAGTGTGAGGAGAGCGGCTTCTTCCGACAGCT 2282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2134 CAGAGACATTCCTCAACAGCAGCAGCACTGGGAACTCCCTGGGCTTCTGCTCGGCAAT 2193
QY 2283 ATAACTTCTTGGAGCAAGACCCCTCTTCTACTGTGCGGCACTC 2325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2194 TCAACTTCTTCCAGAGAGACCCCTCTGCTGGCATCCAGAGACAC 2236

RESULT 7
US-09-918-995-30382
: Sequence 30382, Application US/09918995
: Publication No. US20030073623A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
: TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
: FILE REFERENCE: 20411-756
: CURRENT APPLICATION NUMBER: US/09/918, 995
: CURRENT FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: US/09/235, 076
: PRIOR FILING DATE: 1999-01-20
: NUMBER OF SEQ ID NOS: 38054
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 30382
```

```
LENGTH: 459
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(459)
: OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30382
```

```
Query Match 9.9%; Score 401.4; DB 9; Length 459;
Best Local Similarity 95.0%; Pred. No. 1.5e-109;
Matches 414; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
```

```
QY 821 GGGAGCGGGCCACTGTTGGCTCTACTACAAATCTCGGGGCTGTTCTCTG 880
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 GNGGAATTCGCCACCTAGTTGGCTTACTCAACATCTCGGGGCTGTTCTCTG 82
QY 881 CACCAAGTGGGCTTGGAGTGTCTAGTGAACCAAGGCGCTTACCCCGCGTGGACT 940
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 CACCAAGTGGGCTTGGAGTGTCTAGTGAACCAAGGCGCTTACCCCGCGTGGACT 142
QY 941 ATCCAAAGGTGTCTATCAAGGCGCTACTACAGACGCTGAGCCAGTGGAGCCAG 1000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 ATCCAAAGGTGTCTATCAAGGCGCTACTACAGACGCTGAGCCAGTGGAGCCAG 202
QY 1001 TTTGAGCGCGGCTGTGTAATGTGCTGATCCAGACATGCGACAGTGGTCTGG 1060
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 TTTGAGCGCGGCTGTGTAATGTGCTGATCCAGACAAAGGCAAGTGGTCTGG 262
QY 1061 TCCGTAAGTCCCTGTGGCCCCGGGATCCAGCTCCCGCTACATGTTATCCCAAGC 1120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 TCCCTGAAGTCCCTGTGGCCCCGGGATCCAGCTCCCGCTACATGTTATCCCAAGC 322
QY 1121 CCCGCTTCAAGTGTCCAGGAAGTGAAGTGGCTCTCACTGTGAGCTTCTCTTGGC 1180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 CCCGCTTCAAGTGTCCAGGAAGTGAAGTGGCTCTCACTGTGAGCTTCTCTTGGC 382
QY 1181 CTCGAGCATTTAGTGGCCCAAGATCTTGAACGTTGCTTCCAGGAAGATAGT 1240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 CTCGAGCATTTAGTGGCCCAAGATCTTGAACGTTGCTTCCAGGAAGATAGT 442
QY 1241 TATGATATGAGCTCC 1256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 TATGATATGAGCTCC 458
```

```
RESULT 8
US-09-960-352-6153
: Sequence 6153, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengping
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511, 006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960, 352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 6153
: LENGTH: 444
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 26-LIB34-078-Q1-E1-G9
US-09-960-352-6153
```

```
Query Match 7.8%; Score 315; DB 10; Length 444;
Best Local Similarity 82.8%; Pred. No. 1.2e-83;
Matches 360; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 1293 CAGCAATGACGACCCGCTATGTGATGAGGCTTTGGCATGGGCAAGTACACAGCCGCC 1352
```


Tue May 20 13:18:44 2003

us-10-081-408-1.rnpb

Page 13

Db	136	AGTATGTTCTTCAAGTCGTGTGCAGATGCTCAAAATATGACTTATGTGGGAAATGCT	195
QY	1594	CTTCCACCCCAAGTGGGGCCATAGAAATACAGATTCTATGCGACGGGCTCAATCAGCTGGC	1653
Db	196	CTTCTACTCTCAAAATGGGGCCATTAAATCATCATTCGTCCACAGGGCGCAATTACGTACGC	255
QY	1654	ATTCCCTCTTGTGTCTACTGTGGGAAGTACGGGAACAAGTGTGCAGAGCACACCCTTGGGCAC	1713
Db	256	GTTCTCTGTTGTGTGTGTGCCCAATATACGGAAGACATGTTGGGGAAACACACAGTGTGCC	315
QY	1714	GGTCCACACCACAGGGCCCACTTCAAGGTGATCTGATGTGATGACAGGACTGTGGAGACTG	1773
Db	316	CGTCCAAATAGACAGTCTACTACTACCAACGTGTGATCGGAGCTGTGAGCACTGGCAAACTG	375
QY	1774	GGTCTGGGCGGAGATATGTCTTGTGCCCAAGGCT	1810
Db	376	GGCTGTGCTGATGATATATGCCATTGTGTCTACACAGCT	412

```

RESULT 14
US-09-918-995-35960
: Sequence 35960, Application US/09918995
: Publication No. US20030073623A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
: FROM VARIOUS CDNA LIBRARIES
: FILE REFERENCE: 20411-756
: CURRENT APPLICATION NUMBER: US/09/918, 995
: CURRENT FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: US/09/235,076
: PRIOR FILING DATE: 1999-01-20
: NUMBER OF SEQ ID NOS: 38054
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ. ID NO. 35960
: LENGTH: 405
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-918-995-35960

```

Query Match	5.28;	Score 209.8;	DB 9;	Length 405;
Best Local Similarity	72.78;	Pred. No. 3.9e-52;		
Matches 271; Conservative	0;	Mismatches 102;	Indels 0;	Gaps 0;

Qy	2074	GGAGGAGGAGGCCAGTGCAGCAGACGGCTTTTCATGCAGATGACCCCTGGGCCCCACAGT	2133
Db	13	GGAGGAGGAGTCCACAGAGCAGGTAGCATCTATCACCGAAGATGACATCTGCACACCCACAGT	72
Qy	2134	GCATTTCAGTACTTCATCAACACATGAGACCATTTGCTGGAAAGATTGGTGCCCTGGGT	2193
Db	73	TACCTTGCTACTTCATCAATCAAGATGAAGCCCTTAGGAGAGAGATCGTGCGTTGGGT	132
Qy	2194	GACACGTGCTTTTTCGATATCCACATGCAGAGGACATTTCTTAACACAGTACTGTGGG	2255
Db	133	CACACCCAGCTTCCCTGCACATTCGCCATGCGGAGGACATCCCAACACAGTACTCTGGG	192
Qy	2254	GAACGGCTGGGGCTTTTCCCTCGGACCCCTTACTCTTTGAGAGAGAACCCCTCCTCTTA	2313
Db	193	GAACAGATTTGGCTTTCTTGCTCGGACCCCTATTAACCTTTTGATGAGAGACCCCTCCATCTT	252
Qy	2314	CTCTCCGCACTCCATCTACTTCCGAGGGGACAGATGCTGGGGGCTCTGGAGGTCAACC	2373
Db	253	CTCCCTCTGGCAGTGTCTACTTTGAGAAAGGGGCCAGAGATGCTGGGGCTCTGGAGCATCAATCC	312
Qy	2374	CCTAGCTTGCCCTGCCCCACAGGCTGCTGCTGTGCCCCGACCTCCTGCTTCTCCACGG	2433
Db	313	TGTGGCTGCCTCCCGGACCTGGCAGGCTGTGTCCGGACTTACCCCTTCTCTTACCA	372
Qy	2434	GGGCTTCTCTAC	2446
Db	373	CGGCTTCTAGTAC	385

```

RESULT 15
US-09-960-352-3010
: Sequence 3010, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengding
: APPLICANT: Byatt, John C.
: APPLICANT: Mahalingan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960.352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 3010
: LENGTH: 254
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 13-LIB34-066-Q1-E1-D1
US-09-960-352-3010

```

Query Match	4.5%	Score 183.6;	DB 10;	length 254;
Best Local Similarity	82.7%;	Pred. No. 2e-44;		
Matches 210; Conservative	0;	Mismatches 44;	Indels 0;	Gaps 0;

Qy	1164	GGACCTTCTCCTTTGGGGCTGGAGCAATCATGTGGCCAGAGATCTTTACGTTGGCTTCC	1223
Db	1	GGACTTCTCCTTTGGGGCTGGAGCTTTAGTGTGTCCTTAGGGCTTTAGACGTTCATCTCC	60
Qy	1224	AAGGAGAAAGCATAGTTTATGATTAAGCTCCAGAGGCTTTGGGCATCTATGTGGA	1283
Db	61	AGGGGAAGCGCTGCTTATGATGATCGCTCGAAGGCGCTGTGTCTACAGGTGGGA	120
Qy	1284	ATTCCCCAGCAGCAATACGACCCGGCATGTGATGGAGGCTTTGGCATGGCAGTAGA	1343
Db	121	ATATCCAGCGACATGCTCATCGCTATATGATATGTGGCTTTGGCATGGGTACTTGG	180
Qy	1344	CCAGCGCCCTGACCCGTGGGGGTGAGTCCCTTACTTTGGCACACTACGTGACCTGGCACT	1403
Db	181	CCACACCCCTGATTCGTGGGGGTAGACGCCCTTACTGTGGCACACTACATGAGCACTGGCACT	240
Qy	1404	TCCTTTTGGAGTCC	1417
Db	241	TCGTTGTGGAGTCC	254

Search completed: May 20, 2003, 11:30:46
Job time : 834 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 03:28:53 ; Search time 4937 Seconds

(without alignments)
13252.940 Million cell updates/sec

Title: US-10-081-408-1

Perfect score: 4040
Sequence: 1 gtcctccaccaccctatgc.....cccaaaaaaaaaaaaaa 4040

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estimu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4008.2	99.2	4064	11	BC036368 Homo sapi
2	932.2	23.1	961	9	AL540526
3	817.4	20.2	929	14	BC0706857 AGENCOURT
4	801.2	19.8	887	13	B1818374 603032815
5	780	19.3	1039	9	AL568455
6	751.8	18.6	957	14	BQ878960 AGENCOURT

Result No.	Score	Query Match	Length	DB ID	Description
7	720.4	17.8	764	13	B1489920
8	675.2	16.7	767	12	BC033106
9	634	15.7	644	14	BM973838
10	631.4	15.6	985	14	BQ712543 AGENCOURT
11	622.4	15.4	966	14	BQ720486
12	608.6	15.1	694	12	BQ718238
13	587	14.5	622	14	BM992667
14	567	14.0	1270	14	U25753
15	551.8	13.7	788	13	B1818437
16	549	13.6	703	12	BC721874
17	535.8	13.3	578	13	BM665550
18	527.8	13.1	531	10	BM673388
19	525	13.0	526	14	BM831002
20	508	12.6	655	10	AM853957
21	506	12.5	506	9	A1279210
22	493	12.2	493	9	AL047423
23	482	11.9	482	9	A1129246
24	480.4	11.9	484	9	A1082117
25	469.6	11.6	641	10	BB636655
26	434.4	10.8	436	14	A1042036
27	431	10.7	431	14	BM831093
28	420.8	10.4	570	13	B1681765
29	414.2	10.3	556	13	BM483607
30	413.8	10.2	417	9	AA807904
31	407.8	10.1	412	12	BF433374
32	407	10.1	431	9	A1268118
33	399.2	9.9	541	14	H42560
34	397.2	9.8	424	10	AM131121
35	393.4	9.7	476	9	A1807879
36	384.6	9.5	511	12	BE758045
37	382.2	9.5	410	9	A1871588
38	377.2	9.3	465	9	AA036974
39	376.4	9.3	381	9	A1458535
40	376.4	9.3	426	13	B1013968
41	362.6	9.0	511	14	H43038
42	354.2	8.8	359	10	BE048629
43	354	8.8	354	14	BM823391
44	342.8	8.5	502	10	BE479569
45	333.6	8.3	407	14	W24606

ALIGNMENTS

RESULT 1
BC036368
LOCUS
DEFINITION Homo sapiens, amine oxidase, copper containing 3 (vascular adhesion protein 1), clone IMAGE:4826423, mRNA.
ACCESSION BC036368
VERSION BC036368.1 GI:22137778
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4064)
AUTHORS Strausberg, R.
TITLE Direct Submision
JOURNAL Submitted (05-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NMGRI) & Shiroki Toshitsuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadane@systemsbiology.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 33 Row: n Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA 91: 6806883
This clone has the following problem: frame shifted.

FEATURES

source

Location/Qualifiers

1.4064

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4826423"

/issue_type="Testis"

/clone_id="NIH_MGC_97"

/lab_host="DH105"

/note="Vector: pBluescript"

BASE COUNT 860 a 1177 c 1097 g 930 t

ORIGIN

Query Match 99.2%; Score 4008.2; DB 11; Length 4064;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 4021; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
QY 16 AGTCCAGGATCTGACTACCGGAACTCAGCCAGATCCGGAGGCCGCCACCCGCTC 75
DB 6 AGTACAGGATCTGACTACCGGAACTCAGCCAGATCCGGAGGCCGCCACCCGCTC 65
QY 76 CAGAGGCCAAGAGAGCCCGCTCTGCTGCTGAGAAATCATCTCTCTTGTGTTGA 135
DB 66 CAGAGGCCAAGAGAGCCCGCTCTGCTGCTGAGAAATCATCTCTCTTGTGTTGA 125
QY 136 ATCAGCTGCTCCCTCTTGTGGAATAATGAACAGAGAAATCTCTGCTCTCATCTT 195
DB 126 ATCAGCTGCTCCCTCTTGTGGAATAATGAACAGAGAAATCTCTGCTCTCATCTT 185
QY 196 GCGCGTATCACCATCTTGTGCTGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 255
DB 186 GCGCGTATCACCATCTTGTGCTGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
QY 256 GGGTGAACCCAGGAGCTTCCCATTTGCTCTGTAATCTCCAGTCCAGCCCTTGAC 315
DB 246 GGGTGAACCCAGGAGCTTCCCATTTGCTCTGTAATCTCCAGTCCAGCCCTTGAC 305
QY 316 ACACCCCTGGCCAGAGAGCTGTTGACAGACCTGAGCCGAGAGAGAGAGAGCTGTAT 375
DB 306 ACACCCCTGGCCAGAGAGCTGTTGACAGACCTGAGCCGAGAGAGAGAGAGCTGTAT 365
QY 376 GCGCTTCTGACCCAGGAGGCTGGGAGGAGGCTGTGATGACCCAGGCCGCGCTC 435
DB 366 GCGCTTCTGACCCAGGAGGCTGGGAGGAGGCTGTGATGACCCAGGCCGCGCTC 425
QY 436 GGACAACTGCTCTTCTCACTGAGTGTGACAGCTGCTCCCAAGGCTGAGGCTTGCTCA 495
DB 426 GGACAACTGCTCTTCTCACTGAGTGTGACAGCTGCTCCCAAGGCTGAGGCTTGCTCA 485
QY 496 CTGTGACAGAGGAGAGCCCGCTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 555
DB 486 CTGTGACAGAGGAGAGCCCGCTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 545
QY 556 GCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 615
DB 546 GCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 605
QY 616 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 675
DB 606 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 665
QY 676 GTTCCAGAGAGTCTGAGAGATAGACAGATATCTTCAAGAGAGAGTCCCGCAGGCTTC 735
DB 666 GTTCCAGAGAGTCTGAGAGATAGACAGATATCTTCAAGAGAGAGTCCCGCAGGCTTC 725
```

```
QY 736 TGGGCTTCTCCACCACTGTTGTTCTTACAAGACACCGGAGGAGGAGGAGGAGGAGG 795
DB 726 TGGGCTTCTCCACCACTGTTGTTCTTACAAGACACCGGAGGAGGAGGAGGAGGAGG 785
QY 796 CAGGAGTCCCGGAGTCTGCAATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 855
DB 786 CAGGAGTCCCGGAGTCTGCAATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 845
QY 856 CATCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 915
DB 846 CATCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 905
QY 916 GGGCCTTGAACCTTGGCCGCTGAGATATCAGAAAGGTTTATCAAGAGGCGCTACTACA 975
DB 906 GGGCCTTGAACCTTGGCCGCTGAGATATCAGAAAGGTTTATCAAGAGGCGCTACTACA 965
QY 976 CAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1035
DB 966 CAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1025
QY 1036 AGACAAATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1095
DB 1026 AGACAAATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1085
QY 1096 CCTCTACAGTCTATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1155
DB 1086 CCTCTACAGTCTATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1145
QY 1156 CTCACGTGAGCTTCTCTTGTGCTGAGAGATTCAGTGGCCCAAGATCTTGTGAGCT 1215
DB 1146 CTCACGTGAGCTTCTCTTGTGCTGAGAGATTCAGTGGCCCAAGATCTTGTGAGCT 1205
QY 1216 TCGCTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1275
DB 1206 TCGCTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1265
QY 1276 TGGTGAATTTCCCGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1335
DB 1266 TGGTGAATTTCCCGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1325
QY 1336 CAACTACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1395
DB 1326 CAACTACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1385
QY 1396 CTGGACATCTCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1455
DB 1386 CTGGACATCTCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1445
QY 1456 TGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1515
DB 1446 TGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1505
QY 1516 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1575
DB 1506 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1565
QY 1576 CTATGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1635
DB 1566 CTATGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1624
QY 1636 GGGCTACATCAGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1695
DB 1626 GGGCTACATCAGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1684
QY 1696 AGAGCACACCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1755
DB 1686 AGAGCACACCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1744
QY 1756 AGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1815
DB 1746 AGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1804
QY 1816 CTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1875
```

```
|||||
Db 1805 CTGAGAGCCCTGAGCACCAGCTGCAGAGCTGCAGCTGCCGGAAGCTGGAGATGGA 1864
1876 GGAGCAGAGCGCCCTTCCTCGTGGAGAGGCCACCCCTGCGTACGTATACCTGGGCAGAA 1935
1865 GGAGCAGAGCGCCCTTCCTCGTGGAGAGGCCACCCCTGCGTACGTATACCTGGGCAGAA 1924
1936 CCACAGCAGCAGAGTGGGGTCAACCCCGGGGTACCGCATCCAGATGCTCACTTGTCTGG 1995
1925 CCACAGCAGCAGAGTGGGGTCAACCCCGGGGTACCGCATCCAGATGCTCACTTGTCTGG 1984
1996 AGAGCCCTGCCCCAAAACAGCTCCATGCGAGAGGCTTCACTGCTGGAGAGTACCACT 2055
1985 AGAGCCCTGCCCCAAAACAGCTCCATGCGAGAGGCTTCACTGCTGGAGAGTACCACT 2044
2056 GGCTGTACCCAGCGGAGAGAGAGAGCCAGTACGACAGCGTTTTCATTCAGAAATGA 2115
2045 GGCTGTACCCAGCGGAGAGAGAGAGCCAGTACGACAGCGTTTTCATTCAGAAATGA 2104
2116 CCCTTGGGCCCCCAGCTGTGATTTCACTGACTTCATCAACAATGAGACCATTTGCGAAA 2175
2105 CCCTTGGGCCCCCAGCTGTGATTTCACTGACTTCATCAACAATGAGACCATTTGCGAAA 2164
2176 GGATTTGGTGGCCCTGGGTGACAGCTGTTTCTGCTATTCACCATGAGAGAGCATTTCC 2235
2165 GGATTTGGTGGCCCTGGGTGACAGCTGTTTCTGCTATTCACCATGAGAGAGCATTTCC 2224
2236 TAACACAGTGAAGTGGGGAAGCGGCTGGGCTTCTCCCGACCCCTTAATCTTTGA 2295
2225 TAACACAGTGAAGTGGGGAAGCGGCTGGGCTTCTCCCGACCCCTTAATCTTTGA 2284
2296 CGAAGACCCCTCTTCTACTTGTCCGACTCCATCTACTTCCGAGGGAGCAGAGTCTGG 2355
2285 CGAAGACCCCTCTTCTACTTGTCCGACTCCATCTACTTCCGAGGGAGCAGAGTCTGG 2344
2356 GGCCGTGGAGGTCAACCCCTAGCTGTGCTGCCAGGCTCTGCTGTGCCCGGAGACT 2415
2345 GGCCGTGGAGGTCAACCCCTAGCTGTGCTGCCAGGCTCTGCTGTGCCCGGAGACT 2404
2416 CCCTGACCTTCTCCAGCGGGGCTTCTCTCACACTAGCGGCTCTGCGGATGGGCATGTG 2475
2405 CCCTGACCTTCTCCAGCGGGGCTTCTCTCACACTAGCGGCTCTGCGGATGGGCATGTG 2464
2476 GCCAAGGGCTCCAGGGGCCAGGGGTGTGAGGAGTGGGAGACACTGGGCATGGCCGAG 2535
2465 GCCAAGGGCTCCAGGGGCCAGGGGTGTGAGGAGTGGGAGACACTGGGCATGGCCGAG 2524
2536 CCTGCTCCCTCTTCTCCAGGAGCTCTTCTTCCACTACCTACCCCTCCGATCGG 2595
2525 CCTGCTCCCTCTTCTCCAGGAGCTCTTCTTCCACTACCTACCCCTCCGATCGG 2584
2596 CCTCTGAGCCAGGAGCCCTCTGACCTGTGTGATGCTGCAGACAGGGGACATGAACTTGT 2655
2585 CCTCTGAGCCAGGAGCCCTCTGACCTGTGTGATGCTGCAGACAGGGGACATGAACTTGT 2644
2656 TGATGCCAGCTGTACTGAGTCTCATCCAGAGAGGCCAGGATGGCCAGCTGGAGCGG 2715
2645 TGATGCCAGCTGTACTGAGTCTCATCCAGAGAGGCCAGGATGGCCAGCTGGAGCGG 2704
2716 TGGCCGAGGGGCTCCCTAGATGTTCCCTTGGTGTGCTGTGGGCTTCCGAAATCTTTT 2775
2705 TGGCCGAGGGGCTCCCTAGATGTTCCCTTGGTGTGCTGTGGGCTTCCGAAATCTTTT 2764
2776 AGGCCACCTCCAGAGACTCTAAAAGGGGCTATTCCCTGAGAGCCCGAGAGTAGGGTTGC 2835
2765 AGGCCACCTCCAGAGACTCTAAAAGGGGCTATTCCCTGAGAGCCCGAGAGTAGGGTTGC 2824
2836 CAGTCTGCAAGTGCATAGCTGAGCTGGAAAAGATGCTTGTGCTACATTTCCCTCTATC 2895
2825 CAGTCTGCAAGTGCATAGCTGAGCTGGAAAAGATGCTTGTGCTACATTTCCCTCTATC 2884
2896 CAGTCTGCTCTCTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2955
|||||

|||||
Db 2885 CAGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2944
2956 TGCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3015
2945 TGCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3004
3016 CCCCCTGCTACTCCCCAGCCCTCAGTTCAGTCAAGTTCGCTCTCTCTCTCTCTCTCTCTCT 3075
3005 CCCCCTGCTACTCCCCAGCCCTCAGTTCAGTCAAGTTCGCTCTCTCTCTCTCTCTCTCTCT 3064
3076 GGAAGTCTCAAGGTACAGGGAGCCCTATTCAGAGTGGCCAAATCCCTGTGTCTCTCTCTCT 3135
3065 GGAAGTCTCAAGGTACAGGGAGCCCTATTCAGAGTGGCCAAATCCCTGTGTCTCTCTCTCT 3124
3136 TGTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3195
3125 TGTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3184
3196 TTTCAGGACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3255
3185 TTTCAGGACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3244
3256 GATGTCATCTGACAGAGAGCTCAAGTATAGCATTTTCAAGTATTTGGGGGTGCATG 3315
3245 GATGTCATCTGACAGAGAGCTCAAGTATAGCATTTTCAAGTATTTGGGGGTGCATG 3304
3316 ATATGAAGATGGCCATTTTGTACAGAGGCTCTGTATCTGTCAACAGCCCTGTTGGAG 3375
3305 ATATGAAGATGGCCATTTTGTACAGAGGCTCTGTATCTGTCAACAGCCCTGTTGGAG 3364
3376 GCTGGAGTGAACAAAGGGTGGGCTCAAGAGTGAAGCCAAAGCCCTTCAACTCCA 3435
3365 GCTGGAGTGAACAAAGGGTGGGCTCAAGAGTGAAGCCAAAGCCCTTCAACTCCA 3424
3436 GCCAACCAGCAGAGGGGCTGTCCAAATCAGATTCAGCATGCGAATAGCTGGGCGCTG 3495
3425 GCCAACCAGCAGAGGGGCTGTCCAAATCAGATTCAGCATGCGAATAGCTGGGCGCTG 3484
3496 GGTGAGGTGGGGTCTGGCTAGTGGGAGGGGCTGGCTGGTGGGCGAGGGCCTGGC 3555
3485 GGTGAGGTGGGGTCTGGCTAGTGGGAGGGGCTGGCTGGTGGGCGAGGGCCTGGC 3544
3556 CTGTGCCAGGCTTGGGCTCCATTCCTCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3615
3545 CTGTGCCAGGCTTGGGCTCCATTCCTCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3604
3616 ATGGGACAAAGAAATAGCAGAGATGAGAAACAAACAAACTTTTCTCTAAAGACT 3675
3605 ATGGGACAAAGAAATAGCAGAGATGAGAAACAAACAAACTTTTCTCTAAAGACT 3664
3676 GGTAAATCAATTCATAGACGCTTACAAATACATAGTATGACATTAATAATGT 3735
3665 GGTAAATCAATTCATAGACGCTTACAAATACATAGTATGACATTAATAATGT 3724
3736 ATGCTTATATATACATATATATATATATATATATATATATATATATATATATATATAT 3795
3725 ATGCTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 3784
3796 GTATAGGCACTTCCCATTTTGTGTGAGAAAGAGTGAATATTAACATTAATGTGCT 3855
3785 GTATAGGCACTTCCCATTTTGTGTGAGAAAGAGTGAATATTAACATTAATGTGCT 3844
3856 ATGATGCTATTCAGATTAATGAGTAAACATGATTAATTTTGGAGGGGAACAGTAG 3915
3845 ATGATGCTATTCAGATTAATGAGTAAACATGATTAATTTTGGAGGGGAACAGTAG 3904
3916 GTTGAAGCAGAGAGAGAGAGGCTTAACTTAACCTTATACCTTTTGTACATTTTGAATTT 3975
3905 GTTGAAGCAGAGAGAGAGAGGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 3964
3976 GAACCATGTGACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4035
3965 GAACCATGTGACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4024
|||||
```

OY 4036 AAAA 4040
|||||
Db 4025 AAAA 4029

RESULT 2

AL540526

LOCUS AL540526 961 bp mRNA linear EST 16-FEB-2001
DEFINITION AL540526 LTI_FL002_P11 Homo sapiens cDNA clone CS0E001YN01 5 prime
' mRNA sequence.

ACCESSION AL540526

VERSION AL540526.1 GI:12870759

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1. (bases 1 to 961)

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

CONTACT: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 961

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0E001YN01"

/clone_1bp="LTI_FL002_P11"

/lab_host="DH10B"

/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand

cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-stranded cDNA was digested with Not I

and cloned into the Not I and Eco RV sites of the

PCMVSPORT 6 vector. Library was constructed by life

Technologies. Contact : Feng Liang life Technologies, a

division of Invitrogen 9800 Medical Center Drive Rockville

, Maryland 20850, USA Fax : (1) 301 610 8371 Email :

liang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 184 a 296 c 277 g 199 t 5 others

ORIGIN

Query Match 23.1% Score 932.2; DB 9; Length 961;

Best Local Similarity 99.1% Pred. No. 1.1e-169;

Matches 942; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

OY 1624 ATTCTATGCCAGGGGCTACATCAGCTGGGCTCTCTTTGGTCTACTGGGAAGTACGG 1683

Db 1 ATTCTATGCCAGGGGCTACATCAGCTGGGCTCTCTTTGGTCTACTGGGAAGTACGG 60

OY 1684 GAACCAAGTGTACAGACACCTCTGGGACGTCACACCCACAGCGCCCACTTCAAGT 1743

Db 61 GAACCAAGTGTACAGACACCTCTGGGACGTCACACCCACAGCGCCCACTTCAAGT 120

OY 1744 GGATCTGATGTACAGAGCTGGAGAACTGGGCTTGGGCCAGATATGCTTTTGTCCC 1803

Db 121 GGATCTGATGTACAGAGCTGGAGAACTGGGCTTGGGCCAGATATGCTTTTGTCCC 180

OY 1804 CATGGCTGTGGCCCTGGAGCCCTGAGACACCTGAGAGGCTGACGACCCGGAACCT 1863

Db 181 CATGGCTGTGGCCCTGGAGCCCTGAGACACCTGAGAGGCTGACGACCCGGAACCT 240

OY 1864 GCTGAGATGAGAGAGCAGCGCCCTTCTCTGTGGGAAGCCACCTCTGCTACTCTTA 1923

Db 241 GCTGAGATGAGAGAGCAGCGCCCTTCTCTGTGGGAAGCCACCTCTGCTACTCTTA 300

OY 1924 CCTGGCCGACCAACCAAGTGGGCTACCCCGGGGCTACCCGATTCAGATGCT 1983

Db 301 CCTGGCCGACCAACCAAGTGGGCTACCCCGGGGCTACCCGATTCAGATGCT 360

OY 1984 CAGCTTGTGAGAGCGCGTGGCCCAAAACAGCTCCATGAGAGAGCTTCACTGGGA 2043

Db 361 CAGCTTGTGAGAGCGCGTGGCCCAAAACAGCTCCATGAGAGAGCTTCACTGGGA 420

OY 2044 GAGTACACCTGCTGTGACCCAGCGGAAGAGAGAGCCAGTACAGACAGCTTTT 2103

Db 421 GAGTACACCTGCTGTGACCCAGCGGAAGAGAGAGCCAGTACAGACAGCTTTT 480

OY 2104 CAATCAGATGACCTTGGGCCCCCACTGTGATTTAGTACTTCAATCAATGAGAC 2163

Db 481 CAATCAGATGACCTTGGGCCCCCACTGTGATTTAGTACTTCAATCAATGAGAC 540

OY 2164 CATGCTGGAAGAGATTGGGCGCTGGGAGACAGCTGTTTGTGATATCCACATGC 2223

Db 541 CATGCTGGAAGAGATTGGGCGCTGGGAGACAGCTGTTTGTGATATCCACATGC 600

OY 2224 AGAGACATTCCTTAACACAGTACTGTGGGAGAGGCGGTCTTCTCCGACCTTA 2283

Db 601 AGAGACATTCCTTAACACAGTACTGTGGGAGAGGCGGTCTTCTCCGACCTTA 660

OY 2284 TAACCTTTTGACGAGAACCCCTCTTCTACTGTGCGACTCCATCTACTCGAGGGA 2343

Db 661 TAACCTTTTGACGAGAACCCCTCTTCTACTGTGCGACTCCATCTACTCGAGGGA 720

OY 2344 CCAGATGCTGGGCGCTGGGAGTCAACCCCTAGCTTGGCTGCGCCAGGCTGCTGCTG 2403

Db 721 CCAGATGCTGGGCGCTGGGAGTCAACCCCTAGCTTGGCTGCGCCAGGCTGCTGCTG 780

OY 2404 TGCCCGGACCTCCCTGCTCTTCCACAGGGGCTTCTCTCAACATAGGCGGTCTGGG 2463

Db 781 TGCCCGGACCTCCCTGCTCTTCCACAGGGGCTTCTCTCAACATAGGCGGTCTGGG 840

OY 2464 ATGGGCGATGTGGCCAGAGGCTCCAGAGGCCAGGGTGTGAGGAGTGGGAGAGTGGGCA 2523

Db 841 ATGGGCGATGTGGCCAGAGGCTCCAGAGGCCAGGGTGTGAGGAGTGGGAGAGTGGGCA 900

OY 2524 CTGGGCGGACCTGTGCTCTTCTCTCTGCTGACAGACTCTTTCTTC 2574

Db 901 CTGGGCGGACCTGTGCTCTTCTCTCTGCTGACAGACTCTTTCTTC 950

RESULT 3

LOCUS B0706857 929 bp mRNA linear EST 16-JUL-2002

DEFINITION AGENCOURT_8351612 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6282396

ACCESSION B0706857

VERSION B0706857.1 GI:21845756

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 929)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cga@rsfmail.nih.gov

Tissue Procurement: Dr. Mark Watson

cDNA Library Preparation: Rubin Laboratory

DNA sequencing by: Agencourt Bioscience Corporation

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LCM2476 row: 1 column: 13

High quality sequence stop: 640.

Location/Qualifiers

1. 929

/organism="Homo sapiens"

/db_xref="taxon:9606"

FEATURES

Source

```

/clone="IMAGE:6282396"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Spleen; Vector: pORF1; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT      200 a      271 c      269 g      189 t
ORIGIN

```

```

Query Match      20.2% Score 817.4; DB 14; Length 929;
Best Local Similarity 97.1%; Pred. No. 1.5e-147;
Matches 865; Conservative 0; Mismatches 21; Indels 5; Gaps 3;

QY 1457 GACAGAACACGAGGCTCCCTCGGGGCGACACACTGATCTCTACTGCGACTCTT 1516
DB 14 GAGGAGAACCGAGGCTCCCTCGGGGCGACACACTGATCTCTACTGCGACTCTT 73
QY 1517 GGGGGCTTGGCGAAACGGTGTGCTGCTGATGATGATGATGATGATGATGAT 1576
DB 74 GGGGGCTTGGCGAAACGGTGTGCTGCTGATGATGATGATGATGATGATGATGAT 133
QY 1577 TATGTGTGGGATACGGTCTTCCACCCAGTGGGGCCATAGAAATACGATTTATG 1636
DB 134 TATGTGTGGGATACGGTCTTCCACCCAGTGGGGCCATAGAAATACGATTTATG 193
QY 1637 GGGTACATAGTGGGCTTCTCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1696
DB 194 GGGTACATAGTGGGCTTCTCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 253
QY 1697 GAGCACACCCCTGGGCGACGGTCCACACCCAGCCGCTTCAAGTGTGATGTGTA 1756
DB 254 GAGCACACCCCTGGGCGACGGTCCACACCCAGCCGCTTCAAGTGTGATGTGTA 313
QY 1757 GCAGGACTGGAGACTGGGCTGTGGGCGAGATATGCTTTTCTCCCATGCTGTGCC 1816
DB 314 GCAGGACTGGAGACTGGGCTGTGGGCGAGATATGCTTTTCTCCCATGCTGTGCC 373
QY 1817 TGGAGCCTGAGCAGCAGCTGCGAGAGCTGCGAGTACCAGGAGCTGGAGATGGAG 1876
DB 374 TGGAGCCTGAGCAGCAGCTGCGAGAGCTGCGAGTACCAGGAGCTGGAGATGGAG 433
QY 1877 GAGCAGGCGGCTTCTCTGCTGAGAGCGCCAGCCCTGCTGCTGCTGCTGCTGCTG 1936
DB 434 GAGCAGGCGGCTTCTCTGCTGAGAGCGCCAGCCCTGCTGCTGCTGCTGCTGCTG 493
QY 1937 CACAGACAGAGTGGGGTACCCCGGGGGCTACCGCATCCAGATGCTCAGCTTTGCTGA 1996
DB 494 CACAGACAGAGTGGGGTACCCCGGGGGCTACCGCATCCAGATGCTCAGCTTTGCTGA 553
QY 1997 GAGCCGCTGCCCCAAAACAGCTCCATGCGAGAGGCTTACGCTGGAGAGTACCACTG 2056
DB 554 GAGCCGCTGCCCCAAAACAGCTCCATGCGAGAGGCTTACGCTGGAGAGTACCACTG 613
QY 2057 GCTGTACCCAGCGAGAGAGAGAGCCAGTACGAGCAGCTTTTCAATCAAGATGAC 2116
DB 614 GCTGTACCCAGCGAGAGAGAGAGCCAGTACGAGCAGCTTTTCAATCAAGATGAC 673
QY 2117 CCTTGGGCCCCCAGCTGTGATTTCACTGATTCATCAACATGAGACATTTGCTGAAG 2176
DB 674 CCTTGGGCCCCCAGCTGTGATTTCACTGATTCATCAACATGAGACATTTGCTGAAG 733
QY 2177 GATTGTGGGCTGGGAGAGAGTGTTCCTGATTTCCCATGAGAGAGATTCCT 2236
DB 734 GATTGTGGGCTGGGAGAGAGTGTTCCTGATTTCCCATGAGAGAGATTCCT 793
QY 2237 AACACAGTGAAGTGGGAGAA--CGGCGTGGGCTTCTCTCCGACCTATTAACCTTT-- 2293
DB 794 AACACAGTGAAGTGGGAGAAACGGCGGGGCTTCTCTCCGACCTATTAACCTTTT 853

```

```

QY 2294 -GACCAAGACCCCTCTCTACTGTGCGG-ACTCCATCTACTTCCGAGGG 2342
DB 854 GACCAAGACCCCTCTCTACTGTGCGGAACTCATCTACTTCCAAAGG 904

RESULT 4
LOCUS      B1818374
DEFINITION 603032815p1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174187 5',
            mRNA sequence.
ACCESSION  B1818374
VERSION    B1818374.1 GI:15929130
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 887)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLNL1433 row: m column: 04
            High quality sequence stop: 882.
            Location/Qualifiers
                1..887
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5174187"
                /clone_lib="NIH_MGC_115"
                /lab_host="DH10B"
                /note="Organ: pooled brain, lung, testis; Vector:
                PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                source anonymous pool of 6 male brains, age range 23-27; 1
                male lung, age 27; and 1 male testis, age 69. Library is
                oligo-dt primed and directionally cloned (EcoRV site is
                destroyed upon cloning). Average insert size 1.8 kb,
                insert size range 1-3 kb. Library is normalized and
                enriched for full-length clones and was constructed by C.
                Gruber (Invitrogen). Research genetics tracking code
                021. Note: this is a NIH_MGC Library."

BASE COUNT      252 a      170 c      238 g      227 t
ORIGIN

```

```

Query Match      19.8% Score 801.2; DB 13; Length 887;
Best Local Similarity 98.8%; Pred. No. 2e-144;
Matches 881; Conservative 0; Mismatches 3; Indels 8; Gaps 7;

```

```

QY 3152 TGGGAGTAGAGATTGCTCTACCCCTGCTGGGGCTGGGTGTTTTCAGACAGCTGCT 3211
DB 1 TGGGAGTAGAGATTGCTCTACCCCTGCTGGGGCTGGGTGTTTTCAGACAGCTGCT 60
QY 3212 TCTGTGATTTGTGTGCTGCTGCTATGCTCTCTATAGAGAGAGATGCTATCGTACA 3271
DB 61 TCTGTGATTTGTGTGCTGCTGCTATGCTCTCTATAGAGAGAGATGCTATCGTACA 120
QY 3272 GCAGCAGCTCAAGTATGACATTTCAAGTGATTTGGGGGTGCAATGATATGAAGATGCC 3331
DB 121 GCAGCAGCTCAAGTATGACATTTCAAGTGATTTGGGGGTGCAATGATATGAAGATGCC 180
QY 3332 ATTTTGTACAGGAGGCTCTGTATTTCTGCACAGGCTTTTGGAGGCTGAGTGAACAA 3391
DB 181 A-TTTTGTACAGGAGGCTCTGTATTTCTGCACAGGCTTTTGGAGGCTGAGTGAACAA 239

```

OY	3392	AGGGGTGGGATCAAAAGATAGAGAAGCCAAAAGCCCCCTACAAACTCACCACCCAGCAGAG	3451
Dd	240	AGGGGTGGGATCAAAAGATAGAGAAGCCAAAAGCCCCCTACAAACTCACCACCCAGCAGAG	299
OY	3452	GGGCGTGTCCAATCAACAATTTCAGGCATGCGAATAGCTGGGCCCTTGCGTAGAGTGGGGTCT	3511
Dd	300	GGGCGTGTCCAATCAACAATTTCAGGCATGCGAATAGCTGGGCCCTTGCGTAGAGTGGGGTCT	359
OY	3512	GCCCTAGTGGGAGGGGCGCTGGCCTGGGTGGGGAGGGGCCCTGGCTGGTCCAGGCTTGGG	3571
Dd	360	GCCCTAGTGGGAGGGGCGCTGGCCTGGGTGGGGAGGGGCCCTGGCTGGTCCAGGCTTGGG	418
OY	3572	CCTCCATTCOCATCACTGCTGTCCTCCCTCGAGGCTCTGGATTTGGGATGGGGACAAAGAAAT	3631
Dd	419	CCTCCATTCOCATCACTGCTGTCCTCCCTCGAGGCTCTGGATTTGGGATGGGGACAAAGAAAT	478
OY	3632	ACCAAGAGATGAGAAACAACAAGAAC--TTTTTCTCTTAAGAGACTGGTTAAATCAATCT	3690
Dd	479	ACCAAGAGATGAGAAACAACAAGAACCTTTTTTCTCTTAAGAGACTGGTTAAATCAATCT	538
OY	3691	GATACAGCCTTACAAATACAAATAGTATGCAAGCTAAAAATAATTTGATGCTTTATATACT	3750
Dd	539	GATACAGCCTTACAAATACAAATAGTATGCAAGCTAAAAATAATTTGATGCTTTATATACT	598
OY	3751	AATATGTATAATCTCTCAGGTGAAGAAAGGCAAGCACAGAAATGTATATAGGCACTTCC	3810
Dd	599	AATATGTATAATCTCTCAGGTGAAGAAAGGCAAGCACAGAAATGTATATAGGCACTTCC	658
OY	3811	CATTTGTGTTTCAGAAAGAGAGATATATAACATATATGCTTATATGCTATATCA	3870
Dd	659	CATTTGTGTTTCAGAAAGAGAGATATATAACATATATGCTTATATGCTATATCA	718
OY	3871	GAATAAATGGGTACACACTGATTAATCTTTGGGAGGGGAAACAGTAGTTGAGACAGAGAGA	3930
Dd	719	GAATAAATGGGT-ACACTGATTAATCTTTGGGAGGGGAAACAGTAGTTGAGACAGAGAGA	777
OY	3931	GGGAA-GGGTCTTAACACTTACACCC--TTTTGTACATTTTGAATTTGAACATGTACT	3988
Dd	778	GGGAAAGGGGTCTTAAACACTTACACCCCTTTTGTATATTTTGAATTTGAACCATGTACT	837
OY	3989	GATTTACCTATTCAAATAATAACATTAATGGGCCCAAAAAAAAAAAAAA	4040
Dd	838	GATTTACCTATTCAAATAATAACATTAATGGGCCCAAAAAAAAAAAAAA	887
RESULT 5)	
AL568455/c			
LOCUS	AL568455	1039 bp	mRNA linear EST 16-FEB-2001
DEFINITION	AL568455 LTI_FLOO2_PL1 Homo sapiens CDNA clone CSDBED01YN01 3 prime		
			, mRNA sequence.
ACCESSION	AL568455		
VERSION	AL568455.1	GI:12922811	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1039)		
AUTHORS	L.I.W.B., Gruber,C., Jesse,J. and Polayes,D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES			
source	Location/Qualifiers		
	1..1039		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CSDBED01YN01"		
	/clone_1id="LTI_FLOO2_PL1"		
	/lab_host="DH10B"		
	/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand		

[illegible]

```

Db 192 AGGTGAGTAAATATAAATTAATTCCTTATGCTTATTCCTTATTCAGATAATTAAT 133
Oy 3879 GGGTAACT--GATTACTTTGGGAGGAGCACTAGCTTGGAGACGAGAGGAG 3936
Db 132 GGGTAACTAGATTATTTTGGAGGGGAAACAGTGTGAGACGAGAGGAG 73
Oy 3937 GGTCTT-----ACACTTACACCTTTTGTACATTTTGAATTTGAACCTGTGACTGA 3991
Db 72 GTTTTGTGACTTACTTCTTTTGTGATTTTATTTTATTTTGAATATATGATTKTA 13
Oy 3992 TTACCTATTCNA 4003
Db 12 TTTTATATTAA 1

RESULT 6
B0878960 957 bp mRNA linear EST 16-AUG-2002
LOCUS AGENCOURT_8231141 lupskl_dorsal_root_ganglion Homo sapiens cDNA
DEFINITION clone IMAGE:6182295 5', mRNA sequence.
ACCESSION B0878960.1 GI:22270968
VERSION B0878960.1 GI:22270968
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS NIH-MGC http://mgl.nhl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bhs-riemail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13568 row: a column: 16
High quality sequence stop: 654.
Location/Qualifiers
1..957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:6182295"
/clone_11b="Lupskl_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
directionally cloned using the following adaptors:
5'-TCGACCAAGCGCTCCG-3' and
5'-GACTGATCTAGATCGAGCGGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 168 a 315 c 279 g 194 t 1 others
ORIGIN
Query Match 18.6%; Score 751.8; DB 14; Length 957;
Best Local Similarity 92.5%; Pred. No. 6.8e-135;
Matches 825; Conservative 0; Mismatches 58; Indels 9; Gaps 3;
Oy 29 TGACTACCGGAGACCTCAGCCAGTCGCGGAGCCGCCACCGCTCCAGGACCAACAG 88
Db 1 TGACTACCGGAGACCTCAGCCAGTCGCGGAGCCGCCACCGCTCCAGGACCAACAG 60

```

```

Oy 89 ACCCCCGCTTGTGCTGCGCTGAGAAATACATTCCTCTCTTGTGTAATCAGCTGTCCCT 148
Db 61 ACCCCCGCTTGTGCTGCGCTGAGAAATACATTCCTCTCTTGTGTAATCAGCTGTCCCT 120
Oy 149 CTTCGTGGGAAATGAAACCAAGAACATCTCTGCTCTCTTCTTATTCGCGCTCATACC 208
Db 121 CTTCGTGGGAAATGAAACCAAGAACATCTCTGCTCTCTTCTTATTCGCGCTCATACC 180
Oy 209 ATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 268
Db 181 ATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Oy 269 CAGCTTCCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 328
Db 241 CAGCTTCCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Oy 329 ACCCAGCTTTTTCAGACCTGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 388
Db 301 ACCCAGCTTTTTCAGACCTGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Oy 389 CAGCGGCTGGGCGCAGGCGCTGTGTGATGACAGCCAGCCGCGCCCTCGGACAACTGTGTC 448
Db 361 CAGCGGCTGGGCGCAGGCGCTGTGTGATGACAGCCAGCCGCGCGCTCGGACAACTGTGTC 420
Oy 449 TTCTCAGTGGAGTTCAGAGCTGCTCCCAAGGCTGACAGCCCTGCTCATTGACAGAGGG 508
Db 421 TTCTCAGTGGAGTTCAGAGCTGCTCCCAAGGCTGACAGCCCTGCTCATTGACAGAGGG 480
Oy 509 AGCCCCCACTTCCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 568
Db 481 AGCCCCCACTTCCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Oy 569 AACGTAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 628
Db 541 AACGTAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
Oy 629 GTGAGAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 688
Db 601 GTGAGAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Oy 689 GTGAGATGAGAGAGAGATGATTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 748
Db 661 GTGAGATGAGAGAGAGATGATTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Oy 749 CACTGTGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802
Db 721 CACTGTGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Oy 803 CCCCGTGTCTGCAATCAGGAG--ACCGGCGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 859
Db 781 CCCGTGTGTCTGCAATCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Oy 860 TCGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 911
Db 841 AATCTCGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 892

RESULT 7
B1489920 764 bp mRNA linear EST 28-AUG-2001
LOCUS 603031674F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5172811 5',
DEFINITION mRNA sequence.
ACCESSION B1489920
VERSION B1489920.1 GI:15329148
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 764)
AUTHORS NIH-MGC http://mgl.nhl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

```



```

|||||
Db 121 AAGGTACGGACCCCTAATCAGATGCGCAATCCCTGTGTGCTTCCCTTGTCTGT 180
|||
Qy 3145 TCGTATTGGAGTAGAGTGTCTCTACCCCTGTCTGGGGGTGTTCAGAGC 3204
|||||
Db 181 TCGTATTGGAGTAGAGTGTCTCTACCCCTGTCTGGGGGTGTTCAGAGC 240
|||||
Qy 3205 AGCTGCTTGTGTCATTTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3264
|||||
Db 241 AGCTGCTTGTGTCATTTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
|||||
Qy 3265 CGTGAAGACAGACGCTCACTTCAATTCATGATTTGGGGGCTCAATGATTAAG 3324
|||||
Db 301 CGTGACAGACAGACGCTCACTTCAATTCATGATTTGGGGGCTCAATGATTAAG 360
|||||
Qy 3325 AATGCCATTTTGTACAGAGGCTGTGTATTTCTCAACAGCCTGTTGGAGGCTGAGTG 3384
|||||
Db 361 AATGCCATTTTGTACCA-GGCTCTGTATTTCTCAACAGCCTGTTGGAGGCTGAGTG 419
|||||
Qy 3385 GAACAAAGGGTGGGATCAAGATGAGAAAGCCCACTCAACTCCAGCCACCCAG 3444
|||||
Db 420 GAACAAAGGGTGGGATCAAGATGAGAAAGCCCACTCAACTCCAGCCACCCAG 479
|||||
Qy 3445 CCAGAGGGGGCTGCCAATCATTGATGAGATGAGTGGGGGCTGGGAGGAGTG 3504
|||||
Db 480 CCAGAGGGGGCTGCCAATCATTGATGAGATGAGTGGGGGCTGGGAGGAGTG 539
|||||
Qy 3505 GGGGCTGTGGCTAGTGGGAGGGGCGCTGGCTGGTGGGCA-GGGCTGTGGCTGTCA 3563
|||||
Db 540 GGGGCTGTGGCTAGTGGGAGGGGCGCTGGCTGGTGGGCA-GGGCTGTGGCTGTCA 599
|||||
Qy 3564 GGGTGGGCTGTGGCTAGTGGGAGGGGCGCTGGCTGGTGGGCA-GGGCTGTGGCTGTCA 3623
|||||
Db 600 GGGTGGGCTGTGGCTAGTGGGAGGGGCGCTGGCTGGTGGGCA-GGGCTGTGGCTGTCA 658
|||||
Qy 3624 AAGGAATATACAGAGATGAGAAAGCAAGCAAACTTTTCTTAAGAGCTGTTAAT 3683
|||||
Db 659 AAGGAATATACAGAGATGAGAAAGCAAGCAAACTTTTCTTAAGAGCTGTTAAT 718
|||||
Qy 3684 CAATCTGTATAC 3695
|||||
Db 719 TCATCTGTATAC 730
|||||
RESULT 9
LOCUS BM973838 644 bp mRNA linear EST 21-MAR-2002
DEFINITION UI-CF-ECL1-ach-h-08-0-UI-s1 UI-CF-ECL1 Homo sapiens cDNA clone
ACCESSION BM973838
VERSION BM973838.1 GI:19591429
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 644)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: McCray, PB
McCrays Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

```

```

FEATURES
source
Location/Qualifiers
1..644
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-ECL1-ach-h-08-0-UI"
/clone_1lb="UI-CF-ECL1"
/tissue_type="Lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: pT773-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-ECL1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT773-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(ori)18 tail. The sequence tag for this library is
AAGTCTTAC.
TAG_LIB=UI-CF-ECL1
TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG_SEQ=AAGTCTTAC"
BASE COUNT 155 a 161 c 122 g 205 t 1 others
ORIGIN
Query Match 15.7%; Score 634; DB 14; Length 644;
Best Local Similarity 99.8%; Pred. No. 3.8e-112;
Matches 634; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3406 AGATGAGAGCCAAAGCCCTACACATCAGACCCAGCCAGGAGGGGCTGTCATCA 3465
|||||
Db 636 AGATGAGAGCCAAAGCCCTACACATCAGACCCAGCCAGGAGGGGCTGTCATCA 577
|||||
Qy 3466 CATTGAGCATGCGAATGAGCTGGGCTGGGTAGAGTGGGGCTGTGGCTAGTGGGAG 3525
|||||
Db 576 CATTGAGCATGCGAATGAGCTGGGCTGGGTAGAGTGGGGCTGTGGCTAGTGGGAG 517
|||||
Qy 3526 GGGGCTGGGCTGGGCTGGGAGGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 3585
|||||
Db 516 GGGGCTGGGCTGGGCTGGGAGGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 457
|||||
Qy 3586 CTGCTGTCCCTCTGAGAGTGTGATTTGGGATGGGCAAGAAATGACAGAGATGAGA 3645
|||||
Db 456 CTGCTGTCCCTCTGAGAGTGTGATTTGGGATGGGCAAGAAATGACAGAGATGAGA 397
|||||
Qy 3646 AACACAGAAACTTTTCTTAAGAGCTGTTAATCAATTTGATACAGCTTCA 3705
|||||
Db 396 AACACAGAAACTTTTCTTAAGAGCTGTTAATCAATTTGATACAGCTTCA 337
|||||
Qy 3706 TACAATAGTATGACGTAATAAATATTTGATTTATATATATATATATATCT 3765
|||||
Db 336 TACAATAGTATGACGTAATAAATATTTGATTTATATATATATATATATCT 277
|||||
Qy 3766 TCAGGTGAAAAAGCCAGACAGAAATGTATATAGCCACTTCCATTTGTTCGA 3825
|||||
Db 276 TCAGGTGAAAAAGCCAGACAGAAATGTATATAGCCACTTCCATTTGTTCGA 217
|||||
Qy 3826 AAGAGTAGAATATTAACATATATTTGTTATGCTTTCAGATTAATGGGTAC 3885
|||||
Db 216 AAGAGTAGAATATTAACATATATTTGTTATGCTTTCAGATTAATGGGTAC 157
|||||
Qy 3886 ACTGATTACTTTGGAGGGAGACAGTAGTGGAGAGAGAGAGAGAGGAGGCTTAAAC 3945
|||||

```

Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).

Seq primer: M13 FORWARD
POLYA-Yes.

Db 156 ACTGATTACTTTGGAGGGAGACCAGTAGGTTGAGAGAGAGAGAGAGGCTCTTAAC 97
OY 3946 ACTTACACCCCTTTTGTACATTTTGAATTTTGACAGCTGACTGATTAATCTATTCAAA 4005
|||||
Db 96 ACTTACACCCCTTTTGTACATTTTGAATTTTGACAGCTGACTGATTAATCTATTCAAA 37
OY 4006 TAAACAATTAATGGGCCCAAAAAAAAAAAAAA 4040
|||||
Db 36 TAAACAATTAATGGGCCCAAAAAAAAAAAAAA 2

RESULT 10
B0712543 985 bp mRNA linear EST 16-JUL-2002
LOCUS B0712543
DEFINITION AGENCOURFL 8539807 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6303138
5', mRNA sequence.
ACCESSION B0712543
VERSION B0712543.1 GI:21851442
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 985)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: Dr. Mark Watson
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LNCM2521 row: 1 column: 19
High quality sequence stop: 542.
Location/Qualifiers
1..985

FEATURES
source
1..985
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6303138"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(6). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 211 a 280 c 290 g 204 t

Query Match 15.6% Score 631.4; DB 14; Length 985;
Best Local Similarity 96.2%; Pred. No. 1.2e-111;
Matches 734; Conservative 0; Mismatches 16; Indels 13; Gaps 8;

OY 1457 GAACAGAACAGGGCCCTCCCTGCGGAGACACCACTCAGATCTCTACTGCACTACTTT 1516
|||||
Db 14 GAGAGAACAGGGCCCTCCCTGCGGAGACACCACTCAGATCTCTACTGCACTACTTT 73
OY 1517 GGGGGCTTGGGAAAGGGGCTGGTGTGTCAGATCTATGTCACCTTGGCTCACTATGAC 1576
|||||
Db 74 GGGGGCTTGGGAAAGGGGCTGGTGTGTCAGATCTATGTCACCTTGGCTCACTATGAC 133
OY 1577 TATGTGTGAGATACGCTTCCACCCAGTGGGGCCATAGAAATAGAGTTCTATGCCACG 1636
|||||
Db 134 TATGTGTGAGATACGCTTCCACCCAGTGGGGCCATAGAAATAGAGTTCTATGCCACG 193
OY 1637 GCGTACATCAGCTCGGCAATTCCTTTGGTGTCTACTGGGAAGTACGGGAACCAAGTGCA 1696

Db 194 GCGTACATCAGCTCGGCAATTCCTTTGGTGTCTACTGGGAAGTACGGGAACCAAGTGCA 253
OY 1697 GAGCAGACCCCTGGGCAAGGTCACACACCCAGCCCACTTCAAGGTGATCGATGATGA 1756
|||||
Db 254 GAGCAGACCCCTGGGCAAGGTCACACACCCAGCCCACTTCAAGGTGATCGATGATGA 313
OY 1757 GCAGAGCTGGGAAGTATGGGTGTGGGCGGAGATATGCTTTGTCCTTCATGCTGTGCC 1816
|||||
Db 314 GCAGAGCTGGGAAGTATGGGTGTGGGCGGAGATATGCTTTGTCCTTCATGCTGTGCC 373
OY 1817 TGAAGCCCTGAGACACAGCTCAGAGGCTGCAAGTGAACCCGGAAGTGTCTGAGATGAG 1876
|||||
Db 374 TGAAGCCCTGAGACACAGCTCAGAGGCTGCAAGTGAACCCGGAAGTGTCTGAGATGAG 433
OY 1877 GAGCAGCCGCTTCTCTGTTGGGAAGCCACCCCTGCTACTGTACTTGGCCAGCAAC 1936
|||||
Db 434 GAGCAGCCGCTTCTCTGTTGGGAAGCCACCCCTGCTACTGTACTTGGCCAGCAAC 493
OY 1937 CACAGCAACAGTGGGGTACACCCCGGGGCTACCCGATCCAGATGCTCAGCTTGTGGA 1996
|||||
Db 494 CACAGCAACAGTGGGGTACACCCCGGGGCTACCCGATCCAGATGCTCAGCTTGTGGA 553
OY 1997 GAGCCGCTG-CCCCAAACAGCTCCATGCGAGAGGGCTTCACTGGGAGAGTACCAAGCT 2055
|||||
Db 554 GAGCCGCTGCCCCCAAAACAGCTCCATGCGAGAGGGCTTCACTGGGAGAGTACCAAGCT 613
OY 2056 GCGTGTGACCCAGCGG-AAGGAGAGAG-CCAGTAGCAGACAGCG-TTTTCAATCAGAA 2112
|||||
Db 614 GCGTGTGACCCAGCGGAAAGAGAGAGAGGCCCACTGAGACAGCGTTTTCATCAAAA 673
OY 2113 TGACCCCTTGGGCCCCAC---TGTGATTTCACTGACTT-CATCAACATGAGACATT 2167
|||||
Db 674 TGACCCCTTGGGCCCCACACACGTGGGATTTTCACTGACTTCAATCAACATGAGACATT 733
OY 2168 G-CTGGAAGATTTGG--TGCCCTGGGTGACAGCTGCTTT 2206
|||||
Db 734 GCGTGGAAAGATTTGGGTGGGTGGGCGGAGACGTGGTTT 776

RESULT 11
B0720486 966 bp mRNA linear EST 16-JUL-2002
LOCUS B0720486
DEFINITION AGENCOURFL 8305121 LupsK1_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6193046 5', mRNA sequence.
ACCESSION B0720486
VERSION B0720486.1 GI:21859383
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 966)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM13596 row: a column: 15
High quality sequence stop: 434.
Location/Qualifiers
1..966

FEATURES
source
1..966
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6193046"
/clone_lib="LupsK1_sympathetic_trunk"

```

/sex="male"
/issue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/Note="Vector: pCMV-Sport6 (Life Technologies); site_1:
NotI; site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGGTCG-3' and
5'-GACTAGTCTGATGCGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."
BASE COUNT      153 a      371 c      226 g      216 t
ORIGIN
Query Match      15.4%; Score 622.4; DB 14; Length 966;
Best Local Similarity 94.7%; Pred. No. 6.3e-110;
Matches 655; Conservative 0; Mismatches 36; Indels 1; Gaps 1;
QY 54 TCCGGAGACCCCGCCAGCCGTCAGAGCAAGAGCCCGTCTGCTGCGTGAAGA 113
DB 1 TCCGGAGACCCCGCCAGCCGTCAGAGCAAGAGCCCGTCTGCTGCGTGAAGA 60
QY 114 TACATTGCTCTCTTGGTGGTGAATGAGTCTCTCTTCTGTTGGAAATGAACAGAGA 173
DB 61 TACATTGCTCTCTTGGTGGTGAATGAGTCTCTCTTCTGTTGGAAATGAACAGAGA 120
QY 174 CAATCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 233
DB 121 CAATCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
QY 234 TGTGTGGAGAGGGGTGAGATGGGGGTGAACCCAGCAGCTTCCCATTCCTCTGTAT 293
DB 181 TGTGTGGAGAGGGGTGAGATGGGGGTGAACCCAGCAGCTTCCCATTCCTCTGTAT 240
QY 294 CTCCAGTGGCCAGCCTTGGAGACACCTGGCCAGAGCCAGTGTGGAGACCTGAGCC 353
DB 241 CTCCAGTGGCCAGCCTTGGAGACACCTGGCCAGAGCCAGTGTGGAGACCTGAGCC 300
QY 354 GAGAGAGGTGAGCGGTGATGCGCTTCTGACCCAGGCGCTGGGGCCAGGGCTGAGT 413
DB 301 GAGAGAGGTGAGCGGTGATGCGCTTCTGACCCAGGCGCTGGGGCCAGGGCTGAGT 360
QY 414 ATGACGCCAGGCGCCGCTCGAGCACTGTCTTCTCACTGAGTGTGAGTCTGCTC 473
DB 361 ATGACGCCAGGCGCCGCTCGAGCACTGTCTTCTCACTGAGTGTGAGTCTGCTC 420
QY 474 CCAAGGCTGACAGCCTGCTCACTTGGAGAGGGGAGCCCGCCAGCTGCGCGGAGGAC 533
DB 421 CCAAGGCTGACAGCCTGCTCACTTGGAGAGGGGAGCCCGCCAGCTGCGCGGAGGAC 480
QY 534 TGGGCAATGCTCTTGGAGAGCAACCCAGCCCAAGTGAAGTGAAGTGTGGTGGGG 593
DB 481 TGGGCAATGCTCTTGGAGAGCAACCCAGCCCAAGTGAAGTGAAGTGTGGTGGGG 540
QY 594 CAGTGGCTACCCCTCTCTCACTATGCGGAGCTGACT-GTGGAGGCTGAGGAGCCCTG 652
DB 541 CAGTGGCTACCCCTCTCTATTTGGAGAGCCCACTGTGGAGCCGCAATGAATGCCCTG 600
QY 653 CCGTATACAGAGCCCGCTGCTGTTCAGAGAGTACTGAGACATGAGACATGATCTTC 712
DB 601 CCGTATACAGAGCCCGCTGCTGTTCAGAGAGTACTGAGACATGAGACATGATCTTC 660
QY 713 AACAGAGACTGCGCCAGGCTTCTGGGCTTCT 744
DB 661 CACCGAGACTGCGCCAGGCTTCTGGGAGTTT 692

```

```

DEFINITION 602696217F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4828300 5',
mRNA sequence.
ACCESSION BG18238
VERSION BG18238.1 GI:13997425
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 694)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihai
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM10746 row: a column: 05
High quality sequence stop: 692.
Location/Qualifiers
1. 694
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4828300"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/Note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 124 a 230 c 201 g 139 t
ORIGIN
Query Match      15.1%; Score 608.6; DB 12; Length 694;
Best Local Similarity 97.3%; Pred. No. 3e-107;
Matches 672; Conservative 0; Mismatches 14; Indels 5; Gaps 5;
QY 30 GACTACCGGGAACTGACCAAGTCCGGAGAGCCCGCCAGCCGTCAGAGCCAGAGA 89
DB 6 GACTACCGGGAACTGACCAAGTCCGGAGAGCCCGCCAGCCGTCAGAGCCAGAGA 65
QY 90 GCCCGCGTCTGCTGCGAGATGATGCTCTCTCTGTTGTAATGAGTGTCTCTC 149
DB 66 GCCCGCGTCTGCTGCGAGATGATGCTCTCTCTGTTGTAATGAGTGTCTCTC 125
QY 150 TTGCTGGGAAATGAACAGAAACAAATCTCTGCTCTCATTTGCGCGTATACCA 209
DB 126 TTGCTGGGAAATGAACAGAAACAAATCTCTGCTCTCATTTGCGCGTATACCA 185
QY 210 TCTTTCCCTTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269
DB 186 TCTTTCCCTTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245
QY 270 AGCTTCCCATGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 329
DB 246 AGCTTCCCATGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 304
QY 330 GCCAGCTGTTTGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 389
DB 305 GCCAGCTGTTTGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
QY 390 AGCGCTGGGGCCAGGCGTGTGTGATGACAGCCAGGCGCGGCGCTTGAGACACTGTCT 449

```

```

Db 365 AGCGCTGGGGCCA -GGCTGGATGATGACGCCAGGCCGGCCCTCGGACAACGTGTCT 423
QY 450 TCTCAGTGGATTTGAGCTGCTCCCAAGCGTCGACGCCCTGGCTACTTGGACAGGGGGA 509
Db 424 TCTCAGTGGATTTGAGCTGCTCCCAAGCGTCGACGCCCTGGCTACTTGGACAGGGGGA 483
QY 510 GCCGCCACCTGCGCGGAGGACATGCGCTCTTTGGGAGGCAACCCCGGCCCA 569
Db 484 GCCGCCACCTGCGCGGAGGACATGCGCTCTTTGGGAGGCAACCCCGGCCCA 543
QY 570 A-CGTAGTGAAGTGTGTGTGGGCGACTGCTCACCCTCTACATCGGAGCTGACT 628
Db 544 ACCGTAGTGAAGTGTGTGTGGGCGACTG-CTACACCCCTCTACATCGGAGCTGACT 602
QY 629 GTGAGCGCTATGAGAGCGCCCTGCTCATACCGACCCCGCTGTCTTCCAGAGTAC 688
Db 603 GTGAGCGCTATGAGAGCGCCCTGCTCATACCGACCCCGCTGTCTTCCAGAGTAC 662
QY 689 CTGGA-CATAGACCATGATCTTCAACAGA 718
Db 663 CTGGA-CATAGACCATGATCTTCAACAGA 693

RESULT 13
BM992567/c 622 bp mRNA linear EST 17-JUN-2002
DEFINITION UT-H-DT0-att-f-05-0-UI.s1 NCI-CGAP-DT0 Homo sapiens cDNA clone
IMAGE:5864068 3, mRNA sequence.
ACCESSION BM992567
VERSION BM992567.1 GI:19711956
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 622)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA-Yes.

```

FEATURES source

```

1.622 Location/Qualifiers
/db_organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:5864068"
/clone_id="NCI-CGAP-DT0"
/tissue_type="Metastatic Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
NCI-CGAP-DT0 is a cDNA library containing the following
tissue(s): Metastatic Chondrosarcoma in Lung. The library
was constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dt)18 tail. The sequence tag for this library is

```

```

ACACTTCGG.
TAG-LIB-UI-H-DT0
TAG-TISSUE-Lung metastatic chondrosarcoma
TAG-SEQ-AACTGTTCCG"
BASE COUNT 153 a 151 c 119 g 198 t 1 others
ORIGIN

```

```

Query Match 14.5% Score 587; DB 14; Length 622;
Best Local Similarity 99.4% Pred. No. 4,56-103;
Matches 620; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

```

```

QY 3417 CAAGCCCTCAACTCAGACCAACCAGGAGGAGGCTGTCCAAATCAATTCAGCAT 3416
Db 622 CAAGCCCTCAACTCAGACCAACCAGGAGGAGGCTGTCCAAATCAATTCAGCAT 564
QY 3477 GCGAATGAGCTGGGCGCTGGGTGAAGTGGGGTCTGGCTAGTGGGAGGGCGCTGGCT 3536
Db 563 GCGAATGAGCTGGGCGCTGGGTGAAGTGGGGTCTGGCTAGTGGGAGGGCGCTGGCT 505
QY 3537 GGGTGGGCGAGGCGCTGGCTGGTCCAGGCTTGGGCTCCATTCATCAGCTGTCCCT 3596
Db 504 GGGTGGGCGAGGCGCTGGCTGGTCCAGGCTTGGGCTCCATTCATCAGCTGTCCCT 445
QY 3597 CCGTAGCTGTGATGGGATGGGAGCAAGAAATAGCAAGATGAGAAACACAGAA 3656
Db 444 CCGTAGCTGTGATGGGATGGGAGCAAGAAATAGCAAGATGAGAAACACAGAA 385
QY 3657 CTTTTTCTTAAAGAGCTGTAAATCAATTCGATACAGCTTCAATCAATAGTAT 3716
Db 384 CTTTTTCTTAAAGAGCTGTAAATCAATTCGATACAGCTTCAATCAATAGTAT 325
QY 3717 GCAGCTAAATAATATTGTATGCTTTATATACATATATGTAATATCTTCAAGTGAAA 3776
Db 324 GCAGCTAAATAATATTGTATGCTTTATATACATATATGTAATATCTTCAAGTGAAA 265
QY 3777 AGCGAAGCCACAGAAATGTATAGCGCACTCCCATTTGTTTCAGAAAGAGTAGAA 3836
Db 264 AGCGAAGCCACAGAAATGTATAGCGCACTCCCATTTGTTTCAGAAAGAGTAGAA 205
QY 3837 TATTAACACATTAATGCTTATATGATGCTTATCAAGATTAATAGGTTAACTGATTACTT 3896
Db 204 TATTAACACATTAATGCTTATATGATGCTTATCAAGATTAATAGGTTAACTGATTACTT 146
QY 3897 TTGGAGGGGAAACAGTGTGTGAGGACAGAGAGGAGGCTTTAACTTACACCT 3936
Db 145 TTGGAGGGGAAACAGTGTGTGAGGACAGAGAGGAGGCTTTAACTTACACCT 86
QY 3957 TTGTACATTTTGAATTTTGAACCATGTGACTGATTACATCAAAATTAACAATTA 4016
Db 85 TTGTACATTTTGAATTTTGAACCATGTGACTGATTACATCAAAATTAACAATTA 26
QY 4017 TGGGCCCAAAAAAAAAAAAAA 4040
Db 25 TGGGCCCAAAAAAAAAAAAAA 2

RESULT 14
U25753 1270 bp mRNA linear EST 11-MAR-1996
LOCUS HSU25753 normal female breast tissue Homo sapiens cDNA clone 1308:1
DEFINITION , mRNA sequence.
ACCESSION U25753
VERSION U25753.1 GI:808049
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1270)
AUTHORS Harshman,K., Bell,R., Rosenthal,J., Katcher,H., Miki,Y., Swenson,J.,
Gholami,Z., Frye,C., Ding,W., Dayananth,P., Eddington,K., Norris
,F.H., Bristow,P.K., Phelps,R., Hattler,T., Stone,S., Shaffer,D.,
Bayer,S., Hussey,C., Tran,T., Lai,M., Rostek,K.P.R. Jr., Skolnick

```

TITLE
M.H., Shatuck-Eldens, D. and Kamb, A.
Comparison of the positional cloning methods used to isolate the
JOURNAL
Hum. Mol. Genet. 4 (8), 1259-1266 (1995)
MEDLINE
96090243
COMMENT
Contact: Russell Bell
Myriad Genetics, Inc.
390 Wakara Way, Salt Lake City, UT 84108, USA.

FEATURES
source
1. 1270
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q21: between D17S1321 and D17S1325"
/clone="1308.1"
/clone_lib="normal female breast tissue"
/sex="female"
/tissue_type="normal breast"
/note="Organ: pooled: Vector: lambda gt10: multi-tissue
normalized short-fragment cDNA library used in cDNA
selection experiments per Parmoo et al., 1991, Proc.
Natl. Acad. Sci. USA, 88: 9623-9627."

BASE COUNT
276 a 361 c 359 g 273 t 1 others

ORIGIN

Query Match 14.0%; Score 567; DB 14; Length 1270;
Best Local Similarity 97.7%; Pred. No. 3e-99; Indels 3; Gaps 3;
Matches 606; Conservative 0; Mismatches 11;

432 CCTCGGACAACTGTCTCTCTCAGTGAAGTTCAGTGCCTCCCAAGCTGACCCCTG 491
1 CCTCGGACAACTGTCTCTCTCAGTGAAGTTCAGTGCCTCCCAAGCTGACCCCTG 60
492 CTCACCTGGACAGGGGGA-GCCCCCACCCTGCCGGGAGGACATGCTCTTCTT 550
61 CTCACCTGGACAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
551 GGCAGGACACCCGACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 609
121 GGCAGGACACCCGACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180
610 CTACATGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 669
181 CTACATGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 239
670 CGTCTGTTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 729
240 CATGTTGTTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 299
730 GGTCTGTGGGCTTCTCAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 789
300 GGTCTGTGGGCTTCTCAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 359
790 AATGACACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 849
360 AATGACACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 419
850 CTACACATCTGGGCGTGGGCTTCTCTGACACAGTGAAGTGAAGTGAAGTGA 909
420 CTACACATCTGGGCGTGGGCTTCTCTGACACAGTGAAGTGAAGTGAAGTGA 479
910 CCACAGAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 969
480 CCACAGAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 539
970 CTACAGACCTGGGCGTGGGCGTGGGCTTCTCTGACACAGTGAAGTGAAGTGA 1029
540 CTACAGACCTGGGCGTGGGCGTGGGCTTCTCTGACACAGTGAAGTGAAGTGA 599
1030 GATCCAGACATGGACAG 1049
600 GATCCAGACATGGACAG 619

RESULT 15
B1818437 788 bp mRNA linear EST 04-OCT-2001
LOCUS
DEFINITION
603032714F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174138 5',
mRNA sequence.
B1818437
ACCESSION
B1818437.1 GI:15929307
VERSION
EST.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 788)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LLM11433 row: K column: 03
High quality sequence stop: 755.

FEATURES
source
1. 788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5174138"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH-MGC Library."

BASE COUNT
153 a 256 c 219 g 160 t

ORIGIN

Query Match 13.7%; Score 551.8; DB 13; Length 788;
Best Local Similarity 95.8%; Pred. No. 2.7e-96;
Matches 610; Conservative 0; Mismatches 22; Indels 5; Gaps 4;

1798 TGTCCCATGAGTGTGAGCCCTGAGAGCCCTGAGACAGCTGACAGAGCTGACGCCG 1857
153 TGTCCCATGAGTGTGAGCCCTGAGAGCCCTGAGACAGCTGACAGAGCTGACGCCG 212
1858 GAAGCTGTGTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
213 GAAGCTGTGTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 272
1918 CTTGTACCTGGCCAGCAACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
273 CTTGTACCTGGCCAGCAACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
1978 GATGCTCAGCTTGTCTGGAGAGCCGTGCCCAAAACAGCTCATGGCGAGAGCTTCA 2037
333 GATGCTCAGCTTGTCTGGAGAGCCGTGCCCAAAACAGCTCATGGCGAGAGCTTCA 392
2038 CTGGGAGAGGTATCAGCTGCT--GTGACCCAGCGGAGAGAGAGAGAGAGAGAG 2095
393 CTGGGAGAGGTATCAGCTGCTGTGACCCAGCGGAGAGAGAGAGAGAGAGAGAG 452
2096 AGCGTTTCAATAGATGACCTTGGGAGCCCACTGTGATTTCACTGACTTCATCAAC 2155

Db 453 AGCG-TTTCATCAGAAATGACCTTGGGCCCCCACTGTGATTCAGTACTCATCAAC 511
QY 2156 AATGAGACATTTGCTGG-AAAGATTTGGTGCCCTGGGTGACAGCTGGTTTCTGCATAT 2214
Db 512 AATGAGACCATTTGCTGGCAAGGATTCGGTGCCCTGGGTGACAGCTGGTTTCTGCATAT 571
QY 2215 CCCACATGACAGAGACATTCCTTAACACAGTACTGTGGGGAACGGCCGTGGGCTTCTTCC 2274
Db 572 CCCACATGACAGAGACATTCCTTAACACAGTACTGTGGGGAACGGCCGTGGGCTTCTTCC 631
QY 2275 CCGACCTTAACTTCTTTGACG-AAAGCCCTCTCTTACTGTGCGACTCCATCTACT 2333
Db 632 CCGACCTTAACTTCTTTGACGAAAGACCCCTCTCTACTGTGCGACTCCATCTACT 691
QY 2334 TCCGAGGGGACCAAGATGCTGGGGCTTCGAGTCAACCCCTAGCTTGCCCTGCCCAAG 2393
Db 692 TCCGAGGGGACCAAGATGCTGGGGCTTCGAGTCAACCCCTAGCTTGCCCTGCCCAAG 751
QY 2394 CTGCTGCTGTGCCCGGACCTCCCTGCTTCTCCCA 2430
Db 752 CTGCTGCTGTGCCCGGACCTCCCGGCTTCTCTCCA 788

Search completed: May 20, 2003, 07:58:03
Job time : 4970 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 64.4203 seconds

(without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095

Sequence: 1 MNCKTILVLLILAVITIFAL.....QAAACAPDLPAPFSGHGSN 763

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4095	100.0	763	20	AAV03219
2	2536.5	61.9	729	23	AAU04261
3	457.5	11.2	248	23	ABP41516
4	330	8.1	712	23	ABB91435
5	327	8.0	687	23	ABB02504
6	302.5	7.4	759	23	ABB92082
7	246.5	6.0	300	23	ABB92910
8	214.5	5.2	670	17	AAK94370
9	127	3.1	863	21	AAK42952
10	125.5	3.1	460	23	ABB92909

11	115	2.8	2675	21	AAK07564	Protein encoded by
12	109	2.7	435	14	AAK32989	Rat choline kinase
13	108	2.6	652	23	ABB92424	Herbicidally activ
14	106.5	2.6	446	22	ABG13153	Novel human diagen
15	106	2.6	617	23	AAE23407	Lysine oxidase pro
16	106	2.6	825	22	ABB67566	Drosophila melanog
17	104.5	2.6	594	22	ABG07023	Novel human diagen
18	104.5	2.6	736	23	ABB57339	Mouse ischaemic co
19	104	2.5	1250	22	AAK55914	Amino acid sequenc
20	103	2.5	2224	17	AAW04254	Human factor V. H
21	103	2.5	2224	20	AAK49564	Human lipoprotein
22	102.5	2.5	541	20	AAK01648	Cytochrome P450 en
23	102.5	2.5	541	20	AAK01648	Cytochrome P450 de
24	102	2.5	1194	22	AAK92078	C glutamylum prote
25	101.5	2.5	587	22	ABB52856	Escherichia coli p
26	101	2.5	617	23	AAE23403	Lysine oxidase pro
27	100.5	2.5	718	22	AAU04939	Propionibacterium
28	100.5	2.5	1253	23	AAU10648	Human L1CAM protei
29	100.5	2.5	1257	20	AAW74152	Human L1 cell adhe
30	100.5	2.5	1544	23	AAU97541	Human phospholosl
31	100	2.4	617	22	AAU02204	Trichoderma harzia
32	98.5	2.4	844	22	ABB53083	Escherichia coli p
33	98.5	2.4	897	22	AAK93422	Human polypeptide,
34	98.5	2.4	933	20	AAK26349	Human secreted pro
35	98.5	2.4	933	21	AAK73400	Human secreted pro
36	98.5	2.4	1152	21	AAK94920	Human secreted pro
37	98.5	2.4	1494	23	AAU78460	Mouse beta catenin
38	98	2.4	810	22	ABG04698	Novel human diagen
39	98	2.4	929	20	AAK22189	Timetric murine C3
40	97.5	2.4	308	20	AAK10811	Human secreted pro
41	97.5	2.4	903	21	AAK42926	Human ORF2690
42	97	2.4	545	23	ABB91611	Herbicidally activ
43	96.5	2.4	763	22	AAU01222	Propionibacterium
44	96.5	2.4	816	23	ABB9783	Human polypeptide
45	96.5	2.4	1137	8	AAK71182	Sequence of herpes

ALIGNMENTS

RESULT 1	AAV03219	standard; Protein: 763 AA.
AC	AAV03219;	
XX	21-JUN-1999 (first entry)	
DT		
XX		
DE		
XX		
KW	Human; vascular adhesion protein-1; VAP-1; endothelial cell;	
KW	Lymphocyte; inhibition; amine oxidase.	
OS	Homo sapiens.	
XX		
XX		
FT	Key	Location/Qualifiers
FT	Misc-difference 43	/note= "potential O-glycosylation site"
FT	Misc-difference 47	/note= "potential O-glycosylation site"
FT	Misc-difference 679	/note= "potential O-glycosylation site"
FT	Misc-difference 137	/note= "potential O-glycosylation site"
FT	Misc-difference 137	/note= "potential N-glycosylation site"
FT	Misc-difference 232	/note= "potential N-glycosylation site"
FT	Misc-difference 294	/note= "potential N-glycosylation site"
FT	Misc-difference 592	/note= "potential N-glycosylation site"
FT	Misc-difference 618	/note= "potential N-glycosylation site"
FT	Misc-difference 666	/note= "potential N-glycosylation site"

```
FT Peptide /note= "potential N-glycosylation site"
FT 74..78
FT Peptide /note= "V8 peptide"
FT 170..174
FT Peptide /note= "V8 peptide"
FT 259..262
FT Peptide /note= "V8 peptide"
FT 264..277
FT Peptide /note= "V8 peptide"
FT 323..329
FT Peptide /note= "V8 peptide"
FT 359..361
FT Peptide /note= "V8 peptide"
FT 368..375
FT Peptide /note= "V8 peptide"
FT 384..393
FT Peptide /note= "V8 peptide"
FT 562..566
FT Peptide /note= "V8 peptide"
FT 597..601
FT Peptide /note= "V8 peptide"
FT 624..637
FT Peptide /note= "V8 peptide"
FT 1..25
FT Peptide /note= "N-terminal tryptic peptide"
FT 5..27
FT Domain /note= "transmembrane domain"
FT /note= "transmembrane domain"
XX MO9853049-A1.
XX 26-NOV-1998.
XX 22-MAY-1998: 98MO-F100429.
XX PF
XX 23-MAY-1997: 97US-0862433.
XX PA
XX (BIOT-) BIOTIE THERAPIES LTD.
XX PI
XX Bono P, Jalkanen S, Salmi M, Smith DJ:
XX WPI, 1999-131690/11.
XX DR
XX N-PSDB; AAX28632.
XX
XX Nucleic acid encoding a novel human endothelial cell adhesion
XX protein designated VAP-1 - having an adhesive function and an anline
XX oxidase function useful for manipulating VAP-1 mediated binding of
XX endothelial cells to lymphocytes
XX PT
XX Claim 1; Fig 1; 66pp; English.
XX PS
XX
XX This is the nucleotide sequence encoding the human vascular adhesion
XX protein-1 (VAP-1) used in the method of the invention. The method
XX involves manipulating VAP-1 mediated binding of endothelial cells
XX to lymphocytes which comprises inhibiting the enzymatic activity of
XX CC amine oxidase in endothelial cells, and potentiating the enzymatic
XX CC activity of endothelial cells.
XX
XX Sequence 763 AA:
XX
XX Query Match 100.0%; Score 4095; DB 20; Length 763;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNOKTIVLILAVITTFALVYCVLLVGRGDGGEPSOLPHCPSPVSPSAQFWTHPGQSOLF 60
DB 1 MNOKTIVLILAVITTFALVYCVLLVGRGDGGEPSOLPHCPSPVSPSAQFWTHPGQSOLF 60
QY 61 ADSREELTAVMRELTORLPGLVDAQAARPDNCVPSVVELQLPKKAALAHLDGSGSPP 120
DB 61 ADSREELTAVMRELTORLPGLVDAQAARPDNCVPSVVELQLPKKAALAHLDGSGSPP 120
QY 121 AREALAIVFEGROPQPNVSELVVGPLPHPSYMDVTVERRHGGLPYHRRPVLFQEVLDID 180
DB 121 AREALAIVFEGROPQPNVSELVVGPLPHPSYMDVTVERRHGGLPYHRRPVLFQEVLDID 180
```

```
DB 121 AREALAIVFEGROPQPNVSELVVGPLPHPSYMDVTVERRHGGLPYHRRPVLFQEVLDID 180
QY 181 OMIFNRELPOASGLLHHCCFYKARGRLVTMTAPRGLOSGDBATWFGLYNISGAGFLL 240
DB 181 OMIFNRELPOASGLLHHCCFYKARGRLVTMTAPRGLOSGDBATWFGLYNISGAGFLL 240
QY 241 HHVGLLELVNKKALDPAKMTIOKVFYOGRYDSLAOLEAOFEGGLVNVVLIPIPNCGGSM 300
DB 241 HHVGLLELVNKKALDPAKMTIOKVFYOGRYDSLAOLEAOFEGGLVNVVLIPIPNCGGSM 300
QY 301 SLKSPVPPGAPLLOFPDGRFSSVQSSRVASSLMTFSGFGSRIIDVFQGRRLV 360
DB 301 SLKSPVPPGAPLLOFPDGRFSSVQSSRVASSLMTFSGFGSRIIDVFQGRRLV 360
QY 361 YEISLOEALAIYGNSPRAATTTRYVDGFGMGKRYTTPLRFGVDCPYLATYVDWHFLESQ 420
DB 361 YEISLOEALAIYGNSPRAATTTRYVDGFGMGKRYTTPLRFGVDCPYLATYVDWHFLESQ 420
QY 421 APRTIRPAFCVFQNOGLPLRRHSDLSHYFEGLAETVLVNSMSYLLNYDYVMDTVFH 480
DB 421 APRTIRPAFCVFQNOGLPLRRHSDLSHYFEGLAETVLVNSMSYLLNYDYVMDTVFH 480
QY 481 PSGAIEIRFYATGYISSAFLEFGATGKYGNQVSEHTLGTVTHSAHFVDDVAGLEMMVW 540
DB 481 PSGAIEIRFYATGYISSAFLEFGATGKYGNQVSEHTLGTVTHSAHFVDDVAGLEMMVW 540
QY 541 AEDMVFYPAVWPSPPEHQLOLQVTRKLEMEQAAFLVGSATPRYLYLASNSNKGHP 600
DB 541 AEDMVFYPAVWPSPPEHQLOLQVTRKLEMEQAAFLVGSATPRYLYLASNSNKGHP 600
QY 601 RGYRIOMLSFAGEPLPONSSMARGFSWERYQLAVTORKEEPSSSVFNONDPAAPVDF 660
DB 601 RGYRIOMLSFAGEPLPONSSMARGFSWERYQLAVTORKEEPSSSVFNONDPAAPVDF 660
QY 661 SDFINNETIAGKDLVAVWTAGFLHIPHAEDIPNTVTVGNVGFELRPYNFDEDPSEYSA 720
DB 661 SDFINNETIAGKDLVAVWTAGFLHIPHAEDIPNTVTVGNVGFELRPYNFDEDPSEYSA 720
QY 721 DSIYFRGDODAGACEVNPPLACLPQAAACAPDLPAFSGSGFSHN 763
DB 721 DSIYFRGDODAGACEVNPPLACLPQAAACAPDLPAFSGSGFSHN 763
XX
XX RESULT 2
XX ID AA084261 standard; Protein; 729 AA.
XX AC
XX AA084261;
XX
XX 08-MAY-2002 (first entry)
XX
XX DE Human endometrial cancer related protein, AOC2.
XX
XX KM Human; endometrial cancer; differential expression;
XX DNA microarray; protein microarray.
XX
XX OS Homo sapiens.
XX
XX WO200209573-A2.
XX
XX PD 07-FEB-2002.
XX
XX PF 31-JUL-2001; 2001MO-US24104.
XX
XX PR 31-JUL-2000; 2000US-221735P.
XX
XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
XX
XX PI Mutter GL;
XX
XX DR WPI; 2002-17967/73.
XX
XX DR N-PSDB; ABK35481.
XX
```


CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 248 AA;

Query Match 11.2%; Score 457.5; DB 23; Length 248;
 Best Local Similarity 41.9%; Pred. No. 2,7e-37;
 Matches 83; Conservative 45; Mismatches 67; Indels 3; Gaps 1;

QY 305 PVPGPAPPLQFVPGGPRSVGSRVASSLMTFSGFAGSPRIFFVDFGERTLYEIS 364
 DB 16 PEPHVSGRFLVQPHGPRFRLEGNAVLVGWGFAPRLRSSGLQVLNHFGERIAYEVS 75
 QY 365 LQALAIYIGNSPAAATYTVVVGFGMGKYYTPLTRGVDCPYLATYVDMHFLLESQAPRT 424
 DB 76 VOEAVALYGGHTPAGMOTKYLDVGMGLSVTHELAPGIDCEPTATFLDPFHYDADDPVH 135
 QY 425 INDACVFEQNGGLPLRRHSDLYS---HYEGIAETVIVVBSMTLLVYDVVMDVFERP 481
 DB 136 YRALCLTEMPGVPLRRFRFNSNGGFNFVAGLKGQVLVLTSTVYVYDIWFIEFP 195
 QY 482 SGAEIRFVATGYISSAF 499
 DB 196 NGVMEAKMHATGYVHAXF 213

RESULT 4

ABB91435

ID ABB91435 standard; Protein; 712 AA.

XX ABB91435;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 646.

XX Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN WO200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 CC comprising aligning and comparing nucleic acid or amino acid sequences
 CC from plant with nucleic acid or amino acid sequences from non-plant
 CC organisms -

PS Claim 5; SEQ ID NO 646; 261bp + Sequence Listing; English.

CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

CC Sequence 712 AA;

Query Match 8.1%; Score 330; DB 23; Length 712;
 Best Local Similarity 21.7%; Pred. No. 1.3e-23;
 Matches 172; Conservative 107; Mismatches 346; Indels 166; Gaps 28;

QY 7 LVLLIAVITITFALCYLVLRGCGDG-----EPGSLPHCPSPSPAQWTHPG 55
 DB 9 LFLFFSFLIFATYSWF--GPSGFLGTRVKTIGSNQVH--VDHSLERPHHPL 62
 QY 56 QGQLPADLSREELTAVMRFLTORLPGVDAQAARPSDNCVSEVETOLPPKAAALAHDR 115
 DB 63 DLUTREIR-----VRILSNHDPGESSGAT-----ISMALDEBK-SRVQWKK 109
 QY 116 GSPPARREALVFFGROPQPVNSELVGPPLPHPSYMDYVERHGGPLRYRRPVLEOE 175
 DB 110 GKKLSRRAAVAYVWGGQTHEITVDLDSGRV-----VSDVINRTSGYPI-LTANDVFAAS 163
 QY 176 YLDIDQMIFFRELPAQASGLHHCFCFKHGRNLVMTTPARGLSGDRAT-----WGL 229
 DB 164 QVPLKSLFERNSTI-----EARGVKSDSLCTIPFAGWG- 197
 QY 230 YNNSGAFFLHHVGLLELVNHRKALDPAKWTIQKVFYQGRYDSIAQEAQF---EAGLV 286
 DB 198 -----SDEGRVY-----RVQCFITQGTNYFPRPLEGLVTVTDIDL 236
 QY 287 NVLLPDNCTGGSWSLKSVPVPPGAPPLQF-----YPGSPRSVO-G 327
 DB 237 EVIKIIDKG-----PIPIKASGTEYRFYQVKNPVHMDRIINPISMEQPDPSRFVEDG 289
 QY 328 SRVASSLMTFSGFAGSPRIFFVDFRFGERTLYEISLQALAIYIGNSPAAATYTVV 386
 DB 290 HLYKMANVYHVKADORAGMIT-----SOATVRDSETEGPRSVMTKGPSELEFVYMDP 343
 QY 387 -----GFGMGKYYTPLTRGVDCPYLATYVDMHFLLESQAPRTIRDAQVFEQ 434
 DB 344 EBGWYKGYMDAGELGLPTAMPVPLVNDCCPRNSYIDGVFASPGOKPIVQPMICLFER 403
 QY 435 NOGLPLRRHSDLYSHYFGLAET---VLVYRSMSTLLNTYVWDYVHFGAIEIRFY 490
 DB 404 YAGDISWRHSEILFAN--ADIRESRPKYTLVARMATVSGNYIYIDFWEFOTGLIRYVA 461
 QY 491 ATG-----YISSAFLEGATGKGNVSEHTLGTVHTHSAHPKVDLVAGLENNWAE 542
 DB 462 ASGMLAVKGTPTDNDVDDLDDREDDAGPLISENVIGVYHDHFTFHLMIDIDGPMN---N 517
 QY 543 DMVFVPM--AVPWSPEHQLRLQVTRKLLMEDEQAFLVGSATPRYLYLAS--NHSNKG 598
 DB 518 SLVKVHLEKQRYPTGKSPKSYLKVKYIAKTEKQIKLSLYDPEFIHVPNKRSRVG 577
 QY 599 HPRGRIQLMSTFAGEPLPQNSSMARGFSEWRQQLAVTQKKEEPSSSSVFNQNDPAPYV 658
 DB 578 NPAGRIYVPGNAASILDHDDPQIRGAFTNNQIWTTPRNSRQYAGGLVLIYQSGDDTL 637
 QY 659 D-FSDFINNETIAGKDLVAMVTAAGFLIHPADIPNTVTVGNGVGFLLRPYNFEDPSPF 717
 DB 638 QWMSD--RRRSIENKDIYMTLGFHHVPCQEDPYMPYV--AASSELKPAFESNPIL 693
 QY 718 YSADSIYFRGD 728
 DB 694 GSNP--FFEKD 702

RESULT 5

ABB92504

ID ABB92504 standard; Protein; 687 AA.

XX ABB92504;
 AC 31-MAY-2002 (first entry)
 XX
 DT Herbicidally active polypeptide SEQ ID NO 1715.
 XX
 DE Herbicidal; plant; agriculture; herbicide.
 XX
 KM Arabidopsis thaliana.
 XX
 OS WO200210210-A2.
 XX
 PN 07-FEB-2002.
 XX
 PD 28-AUG-2001; 2001WO-EP09892.
 XX
 PF 28-AUG-2001; 2001WO-EP09892.
 XX
 PR (FARB) BAYER AG.
 XX
 PA Tietjen K, Weidler M;
 XX
 PI WPI: 2002-269010/31.
 DR
 XX Identifying plant target proteins for herbicidally active compounds,
 XX comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 PS Claim 5; SEQ ID NO 1715; 261pp + Sequence Listing; English.
 XX
 CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.
 CC
 CC Sequence 687 AA;
 SQ
 Query Match 8.0%; Score 327; DB 23; Length 687;
 Best Local Similarity 21.6%; Pred. No. 2,4e-23;
 Matches 171; Conservative 113; Mismatches 314; Indels 194; Gaps 31;

DB 281 HIRPDQACMI-----SQATVRDSKTEGARSVMKGFASSELFPNMDGEGYSAKM 334
 QY 386 DGG-FGKGYTPTLRGVDPCPYLATYVDVHFLLESQAPKTIKDAFCVFEONOGILPLRRH 444
 DB 335 DAGEFTGLGSSMLVPLNDPPRNAIYIDGFFASPEGPIIOPNNICLFERYADTSRHS 394
 QY 445 SDLYSHYFGL-----AETLVVRSMTLLNDYMDVTFHPSGAIEIRFATG----- 493
 DB 395 EIL-----LPGVDIREBRAKTLVARMACSYGNVDYIDFMEFQMDGVIRVTVAAAGMLVK 450
 QY 494 ----YISSAFLEPGATGKGNQVSEHTLCTVTHSHAFVVDLDVAGLEMMVAEDVFPMA 550
 DB 451 GTAENVEDDEGEKEDDGGPISENVIVVDHFIIFHLDMDIDGSAN-----NSFVKVHLE 506
 QY 551 VPMSPEHQLOR---LQVTRKLEMEEOAAFLVGSATPRYLTLAS-NHSNKGHPRGYRIQ 606
 DB 507 KQRLPGESESRKSLKLYKKYAKTEKDAQIKMSLYDEYEFHLVAPNLSRLGNAGYKLV 566
 QY 607 MLSFAGEPLPONSSMARGFSEWERYOLAVTORKEEPESSSVFNQNDPMA----- 655
 DB 567 PGNAASILDHDDPPQMRGAFNNQIWYTR-----YNSQEWAGGLMYQSRGE 615
 QY 656 PTVD-FSDFINNETIAGKDLVAMWTAGFLHHPHADIPNTYNGNGVGFILRPYFDED 714
 DB 616 DTLOVMSD--RDRSIEKKDILWYTLGFHHVPCQEDPEPMTIAS--SEELKPVNFESN 671
 QY 715 P-----SFYSAD 721
 DB 672 PVLGISPFEEKD 683

RESULT 6
 ABB92082
 ID ABB92082 standard; Protein: 759 AA.
 XX
 XX ABB92082;
 AC
 XX 31-MAY-2002 (first entry)
 DT
 XX Herbicidally active polypeptide SEQ ID NO 1293.
 DE
 XX Herbicidal; plant; agriculture; herbicide.
 KM
 XX Arabidopsis thaliana.
 OS
 XX WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-EP09892.
 XX
 PR 28-AUG-2001; 2001WO-EP09892.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Tietjen K, Weidler M;
 XX
 PD WPI: 2002-269010/31.
 XX
 PD Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 PS Claim 5; SEQ ID NO 1293; 261pp + Sequence Listing; English.
 XX
 CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant

KM butyl amine; benzylamine; oxidative deamination;
 KW recombinant production.
 XX Aspergillus niger.
 XX JP08070872-A.
 XX 19-MAR-1996.
 XX 05-SEP-1994; 94JP-0211526.
 XX 05-SEP-1994; 94JP-0211526.
 XX (TOKU) TOKUYAMA SODA KK.
 XX MPI: 1996-203152/21.
 DR N-PSDB; AAT14174.
 PT Gene encoding amine oxidase - used in formation of hydrogen peroxide
 PT from butyl amine, or benzylamine by oxidative deamination
 PS Claim 3; Pages 9-11; 12pp; Japanese.
 CC The present sequence is the A. niger strain M-62 amine oxidase
 CC (AO), which is used in the formation of hydrogen peroxide from
 CC butyl amine, or benzylamine by oxidative deamination. The AO gene,
 CC the sequence of which was determined by sequencing a restriction
 CC enzyme map of an AO gene contg. cDNA, can be used for the
 CC recombinant prodn. of AO.
 CC
 SQ Sequence 670 AA:
 Query Match 5.2%; Score 214.5; DB 17; Length 670;
 Best Local Similarity 22.3%; Pred. No. 5,6e-12;
 Matches 168; Conservative 86; Mismatches 301; Indels 197; Gaps 37;
 QY 53 HGGQSLPRLDLSRELTAVMRPLTORLGGVLAQAARSDNVFVELQPKA----- 107
 DB 4 HP-----LAILSEETNIAKNVLAQHPTVIDERE-----IYSEPKAOLLER 48
 QY 108 AALAH---LDRGSPPREALAA---IVFGRGP-----QPNVSELVVGPLPHPS 150
 DB 49 LALEHSGRLSPSPRRPRALACQDYVIGNDRTSFEEVSVDVGTQRQVRVVGKHEHAS 108
 QY 151 YMR---DVTVER-HGGPRLPHRRPLVLEQYL-DIDQMIFNELPQA-----SGT--- 194
 DB 109 LTLSEFDTLVERCFASP-----LFOKALADFD-----LPEGEFVIEEYCGIDYV 153
 QY 195 -----LHNCCEFKHGRN-----LVTMTAPRGOSGDRATWFGLYNISG 235
 DB 154 EERKRYFOGLCFATYDRKKNPNANFYSLPLLPVMDGAPGSHGSPSR-----HKCKG 208
 QY 236 AGFFLHHVGLLELVNKHAKLDPRAMTIQKVEYQGRYVDSLQLEAQAAGLVNVLIPDNG 295
 DB 209 EGIETDQTFKRDITGHCKDSP-----IYVELL 234
 QY 296 TGGSWSLKSPVPPGAPRPILOFYQGRPSV-QGSRVASSIMTSPGL-----GAFSGPRIF 350
 DB 235 PGGTREDLKLPLN-----VVOEGSPSFRTESLSIVEMQKMRFRVAFNPRGCHSOTSW 287
 QY 351 DVAFQGERLVYEISLOEALAIYGNSPAAMTTRVYD-GGGMCKYTTPLRLRGVDCPLLAT 409
 DB 288 ---YDRNSVLRLSVSMTVPYADPRPFHRKQAFDGGDGGGMANNLSIGCCCLAVIK 344
 QY 410 YDVMHFLLESQAPKTIIRDAFCVFEQNOGLPLRHHSDLYHYEGSLAETV-----LVVRSM 465
 DB 345 YEPNAVMTGADGSAKKMNAICLHEDONGIGK-----HSNMRKGRAVYTRHRELVAQFI 398
 QY 466 STLLNTDYVMDVYFHPGSAIEIRFYATGYISSAFLGATSKY-----GNQVSE-----HTLG 517
 DB 399 ITLANVEYIFAUKPDGSGGIGTGRAVACHYLE-----RQHRCWQCGQRYVQRORQWBRVG 451
 QY 518 ---IVHNSAHFKVDLDVAGLENNVNAEDMVFYPMVPSPEHQLOLQVTRKLEHDEQ 574

DB 452 PEPPAHFLCAHRP---GYGPNNSVOVESHSHPVPAANATVPNGNRYKVVT-----ETMR 503
 QY 575 AAFVGSATP---RYLYLASNH-----SNKMGHPRGYRIOMLSFAGEPLPQNSSNARGPS 626
 DB 504 AGFF---DAPELNTVTKVMNPHKKNPISOK---PYGKFIPLATQRLRLADPNSIOARRAQ 558
 QY 627 WERYOLAVTORKEEPPSSSVFN-QNDPAPAPVYDSDFI-NNETLAKGLVAMVATAGFLH 684
 DB 559 FAQHVMVWTKYKRDGELYAGGRYTLQSQEIEGV--SDAVKRDSDVDTDVVWSTFGITH 616
 QY 685 IPHAEDIPVTVVNGVGFFLRPNYFDEDP 716
 DB 617 NPREWDP--VMPVELFQMLMRPADEFTANPS 646
 RESULT 9
 AAB42952
 ID AAB42952 standard; Protein; 863 AA.
 AC AAB42952;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORF2716 polypeptide sequence SEQ ID NO:5432.
 XX
 KW Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 KW vulnerability; antiparkinsonian; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic; antihypertensive;
 KW antiviral; antibacterial; antifungal; antipneumatic; antihypertensive;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; nocurnal haemoglobinuria; burn; wound;
 KW thrombosis; contraceptive.
 KW Homo sapiens.
 OS
 XX
 PN MO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinketsu RA, Leach M;
 XX
 DR MPI: 2000-602362/57.
 DR N-PSDB; AAC77161.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 4615-4617; 5507pp; English.
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerability;
 CC antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypertoid; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 863 AA;

Query Match 3.1%; Score 127; DB 21; Length 863;
 Best local Similarity 22.3%; Pred. No. 0.0061;

Matches 100; Conservative 34; Mismatches 126; Indels 188; Gaps 22;

QY 30 GDGGEPSQLPHCPSPVSP--SAQPTNHP---GOSQLFADLSREELTAVMRELTQRLGPG 83
 DB 388 GAGGAPASATPTPALSPGRSLRPHILPLILRGAPAPLTDACQCMCKLRGAQGPLGP-- 445
 QY 84 VDAQAARSPDNCVSVELQLPKAAALAHLDRCSPPARP-----ALAIYFGQOPQPNV 138
 DB 446 -----DMSPLRPPPLSLRPGAGAPPPKPNPRLMALALERRAQVAEQ 491
 QY 139 SELVGPDLPHPS--YMRDVTVERHG-----GLPYHRRPV 172
 DB 492 SQDCCGGTTPPASOSPFHRSLSLEVGGERPLTSGSGPPNSLAHGCANVPGPPLPR--- 548
 QY 173 FQEVLDIDQMFNRELPOASGLLHCCFYHNRGNLVTMTAPRGLOSGDRATWFGLYYN 232
 DB 549 -----QOSGSLT-----RSQRPMTGSRGLRGP----- 572
 QY 233 ISGAGFLHNVGELLVNHKALDPAWTKVFGRYYSOLAQLEAFGLVNVYLIP 292
 DB 573 -----AOVSAQLRAG----- 582
 QY 293 DNGTGS-----WSLKSPVP-PG---PA-----PPL-----QFYPOGPRFSVQ 326
 DB 583 -----GGGRDAPRAAAGSPSCVSPQVPTPPGFSPAPRCLPFLGVPRGLVPLRP--SFQ 637
 QY 327 GSRVASSLMTFSGLGAFSGPRITDVFRQGERLYVEISLOBALIYGCN---SP-AAVTT 382
 DB 638 PSSPA-PVWRSSTLG-----PPAPLD---RGENTLYEIGASEGSPYSGPTRSMSPFRSKMP 688
 QY 383 RYVDGFGMGKYYTPTLRGVDCPYLATY 410
 DB 689 DRUNASTGMGLQSPPLHRSPD--FLLSY 714

RESULT 10

ABB92909

ID ABB92909 standard; Protein; 460 AA.

XX ABB92909;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 2120.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX W0200210210-A2.

XX 07-FEB-2002.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PF 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidler M;

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,

XX comprising aligning and comparing nucleic acid or amino acid sequences

XX from plant with nucleic acid or amino acid sequences from non-plant

XX organisms -

XX Claim 5; SEQ ID NO 2120; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins

XX (ABB90790-ABB94016) for herbicidally active compounds, comprising

XX aligning and comparing nucleic acid or amino acid sequences from plant

XX with nucleic acid or amino acid sequences from non-plant organisms using

XX suitable search parameters, where plant sequences having an E-value

XX greater by a factor of 3 than the E-value of most similar non-plant

XX sequences are selected. The polypeptides or nucleic acids encoding them

XX are useful for identifying modulators. The identified modulators are

XX useful as herbicides.

XX Sequence 460 AA;

XX Query Match 3.1%; Score 125.5; DB 23; Length 460;

XX Best local Similarity 20.7%; Pred. No. 0.003;

XX Matches 101; Conservative 49; Mismatches 186; Indels 153; Gaps 22;

QY 31 DGGEPSQLPHCPSPVS-----PSAOPWTHPGOSQL---FADLSREELTAVM 72
 DB 39 DCRKSSSPFCASNFLENKNTREPSKNIPTKPKNDHVSDTYRNPDLPTVMEINKVR 98
 QY 73 RFLQRLGPGLVDAQAARSPDNCVSVELQLPKAAALAHLDRCSPPARPAPALAIYFGR 132
 DB 99 SILSSH-----ALFASRVPHLNLSVLEEDPK-NLYROMEKGDQLPPKASVYARVG 150
 QY 133 QOPNVSELVVGRLPHPSYMDVTVERHGGLPYHRRPVLFQEYLD-----IDQMFNR 186
 DB 151 N-----SHLIVLDL---STSR---VDQADSPVPSGVPYITSEEMDSAAAPFENADFN 199
 QY 187 ELPOASGLLHCCFYHNRGNLVTMTAPRGLOSGDRATWFG-----LYYNI 233
 DB 200 TI-----NSRGVNLTDVICIP--ISSG---WFGNMDTKRYTKIQCFSTQ 240
 QY 234 SGAGFLHNV-GLELLVNHKALDPAWTKVFGRYYSOLAQLEAFGLVNVYLIP 292
 DB 241 DTPNFYMRPIEGTLT-----FDLDTKRILEIT 268
 QY 293 DNGTGSWSLKSPVPGPAPLPD-----YPQGRFVSOGSRVAS-S 333
 DB 269 DTGQ-----SIPFGSTNDYRSTLPNHDKTRPLMTISLEQRGGSFVIEDHLLKWA 322
 QY 334 LWTFSF-----GLGAPSGPRITDVFRQGER-LYVEISLOBALIYGCNSPAAATTRYVD 386
 DB 323 NMEFHLPDPRAAGL-IMSQVKIHDPDQETREVMYKGFVSELEFVYMDPSAATFVKTYMD 381
 QY 387 GG-FGMGKYYTPTLRGV-----DCPYLATYDMHFLESQAPKTIROAFCYFEQ 434
 DB 382 AGEYGFGLQAMPLEPVAEPKNTLGTPLNDCPRNAVYMDGFAAADGTPYVRENMICVFE 441
 QY 435 NOGLPLRRH 443
 DB 442 YAGDIAMRH 450

RESULT 11

AA07564

ID AA07564 standard; Protein; 2675 AA.

```

Query Match          2.8%   Score 115; DB 21; Length 2675;
Best Local Similarity 18.9%; Pred. No. 0.65;
Matches 172; Conservative 107; Mismatches 321; Indels 310; Gaps 42.

QY      48 AQPWTHPGOSQLFADLSREELTVANRFLTORLSPGLVDAQAQR---PSDCNCFSEYELDLP 104
       :|::||::||::||::||::||::||::||::||::||::||
Db     1329 ARPCGHNDPRLVRRIAREAVTTV-HFVPSMLTFPELTARCTTRLRALRVCSGEELP 1387

QY      105 PKAA--ALANDLR---GSPRPAREALAIYFCG-ROROPVSELVNGPLPHBPMRDYLP 158
       :|::||::||::||::||::||::||::||::||::||::||
Db     1388 AAAYNRAGLLDARLYNLGYPTAEADVDTAMCPRRER-----GPVIGIDPIANTTAE 1440

QY      159 RHGG---PLYNHRRPVLEOEYLIDIQMIRNELPQASGLHHSCFKHGRNMLVTMTTPAR 215
       :|::||::||::||::||::||::||::||::||::||::||
Db     1441 VIDGRLRPR---RPPRGELYL-----GGACLAN-GYNHDPALTAARLPARP 1483

QY      216 RG---LOSQDRATWEGLYLVNISGAGFFLHNHGVELLVNKAKLDAPRMWTIOKFVOGORYUD 272
       :|::||::||::||::||::||::||::||::||::||::||
Db     1484 GGGRRYRTGDL-----VRRADGALIVFRRTDDQVAIGGIINRF-----GEVAN 1527

QY      273 SLAQLDEAFGLVVNVVLLIPDGTSMSKSLSPVPGP----- 310
       :|::||::||::||::||::||::||::||::||::||::||

```

Db	1528	ALBAL	-----PGADAAYVHDGRLAAVAADVPVGPAPADAALBDALRRRLPGHLVPAAL	1582
QY	311	-----	---APPLQFYPOGPRPFVSQGS--RAASSLMTSTFSGLAFSG	346
Db	1583	TLLDRPLPLPAGLBYDRRALPHPSAPP	---PBGGRPPPTTGTERLVARWV	1627
QY	347	PRIFDVRFOGGERLYVE--ISLQELALVYGNSPNAATTTTYVGDFGGMKGYTTPLTRGVDCP	405	
Db	1628	-----AERLGREVVAVDNDQDFSSLGQDSYRALGCVTAALBAALGRLPTVTDLLRLPTVA	1678	
QY	406	YLATYVDWMH-----	---FLT--ESQAPKTIKIDAFVCFEONQGLPLRRHSDLYS	449
Db	1679	ALAHADDERADRDRPARQETPRPGPRFALCERPAAGVPLLEDAVPRMSMAQVAVLFRDHNPGE	1738	
QY	450	HYFGGLA---ETVLVBSMSTLLN-YDYV---MDTYFHPBSGA-----	TEIRFY 490	
Db	1739	VYVTSVAVSVPDLDRTRLAAAVDRLLDLRHAYLRSSEFDLVSHPEPTQLVTHLPTPLEV---	1795	
QY	491	ATGVYSAPFLFG-----ATGKYGNQVSHETLGTVHTH---SAHFKYDLDAVAGLENNYMA	541	
Db	1796	-----VESSDPAGFDAMLHAERKRPDLVGTGTPLAFTTAHDACAAGRLTVSSFALDGKCYA	1851	
QY	542	EDMWYV-----	---PAAPV-----MSPE-----HQL 559	
Db	1852	TVLTTELLROYMSALRGCAPLSLPAPASVREFVALERAAQHDPANHEFPRTTELACARPHL	1911	
QY	560	QRLQVT-----	---RKLEMEEOAA-----FLVG-----SATP 584	
Db	1912	PRRPVPRPGDGIROHGHVYVPEDTVAKGLSALAGELGVLKHLVGLVHLRVYRALSDP	1971	
QY	585	RILYTLAHSNHNKNGH-----	---PRGYRIQLSPAGEPLPLQNSMARGSEWERYQ 631	
Db	1972	DVITAVETHRLERLHHODRVLGVFNNILPLRQVQDGSMAADLARAHAHAEAATGETRRPY	2031	
QY	632	LAVTORKEEPPSSSSVNONDPAPTYDESDFINNETIGKDLVAVMTAGFLHIPHAEDI	691	
Db	2032	LAQAGR---DHGAAGLFD-----	---TLFVTHFLHYALALDDGMA---VSDLRADQTYV 2079	
QY	692	PNT-----VTVNGVGEFLR--PYNFEDDEPSEFYSADSIYFRGDODAGACEVNFACLAP	743	
Db	2080	PLTAHFVWDATDGGGLRLLESDPRREFPDEQVAEFPA---YYRR-----	AL 2122	
QY	744	QAACARDLP 753		
Db	2123	RAAADAPHRP 2132		
RESULT 12				
AAR32999				
ID	AAR32999 standard; Protein: 435 AA.			
XX				
XX	AAR32999;			
XX				
DT	14-JUN-1993 (first entry)			
XX				
DE	Rat choline kinase.			
XX				
KW	CK; epitope; tumour; diagnostic; antibodies; distribution; liver;			
KW	cranial nerve disease.			
XX				
OS	Rattus rattus.			
XX				
FH	Key			
FT	Peptide 139..225			
FT	Peptide 278..327			
FT	Peptide 328..360			
FT	Peptide 398..412			
XX				
PN	JP05015367-A.			
XX				
PD	26-JAN-1993.			
XX				
PF	17-SEP-1991; 91JP-0262516.			

XX 14-SEP-1990; 90JP-0242528.

XX (MITK) MITSUI TOATSU CHEM INC.

XX MPI; 1993-070176/09.

DR N-PSDB; AA037687.

XX Rat choline kinase having specified DNA sequence - useful for
PT diagnosing cranial nerve diseases, liver diseases, tumours, etc.

PS Claim 1; Page 21; 26pp: Japanese.

XX The choline kinase (CK) gene may be obtd. from rat brain, human
CC brain, human liver, chicken liver, egg, soy bean, etc. The rat CK
CC gene is shown, and can be isolated by PCR using oligonucleotides
CC corresp. to partial CK sequences as primers. The oligonucleotides
CC or the amplified DNA fragments, can then be used as probes to detect
CC CK coding sequences. Rat CK may be purified by MMA-agarose affinity
CC chromatography. A large amt. of CK may be produced and antibodies
CC against CK may be obtd. for use in detecting the distribution of CK
CC in tissues or cells, and the diagnosis of cranial nerve disease,
CC liver disease and tumours.

XX Sequence 435 AA;

Query Match 2.7%; Score 109; DB 14; Length 435;

Best Local Similarity 21.5%; Pred. No. 0.13; Mismatches 150; Indels 98; Gaps 14;

Matches 79; Conservative 41; Mismatches 150; Indels 98; Gaps 14;

QY 268 GRVYDSLQLEAFACGLVNVYLIPDNGTGGWSLKSPPVPPAPLQFYQGFPSVQG 327

DB 33 GQQRDAAGELSKOLGSRQPLALPPPPP-----PLPLPPPPPLADEQAPR----- 82

QY 328 SRVASSLWTSFGLGAFSGPR-----IFDVRFQGERLYVEISLQDALIYCGNSPAATT 382

DB 83 TRRAVYLMCKEFLPGAWRGLEDOFHISVIRGRLNMLFCQSLPDSIASV-GDEPRKYL 141

QY 383 RYVDGFGMGKTYTPLTRGVDCPYLATYVDWHFLLSQAPKRTIDAFVFEQ---NOGLP 439

DB 142 RL-----YGAILMKGAEMVLESVMFAILERSLGR-----LVGIFQGLRQPIR 188

QY 440 LRRHSD--LYSHYFGGLAEVLYVRSNSTLNDYVDVYFHFSGAIEIRFYATGYISS 497

DB 189 SRRIDTELCLPDISABIEAKMATFHKMKPFNKPKV----- 226

QY 498 AFLFGAGKXGNOVSEHTLGTVHTSAHFKYVDLVAGLENNVMAEDVFMAYPWSPEH 557

DB 227 --LGTMEKTYLNVLR-----LFSREARVQQLHK-----FLSYNLPLELEN 266

QY 558 QLORLQVTRK-----LLE---MEQAAFLVGSATPRYLTLA---SNHSN 595

DB 267 LRSLLQYTRSPVYVCHNDCCGNNLLLEGGQNSERKQKMLIDFEYSSTVYNGFDIGNHFC 326

QY 596 KWGHPRGY 603

DB 327 EMMYDYTY 334

RESULT 13

ABB92424 standard; Protein: 652 AA.

AC ABB92424;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1635.

XX Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN MO200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR MPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.

PS Claim 5; SEQ ID NO 1635; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

SQ Sequence 652 AA;

Query Match 2.6%; Score 108; DB 23; Length 652;

Best Local Similarity 20.1%; Pred. No. 0.32; Mismatches 190; Indels 240; Gaps 28;

Matches 122; Conservative 56; Mismatches 190; Indels 240; Gaps 28;

QY 11 ILAVTIFALVCVLLVGRGDGGEPSQLPHCPSPSAOPWTHPGQSLFADLSREELTA 70

DB 151 ILVITLTLCLLCKKKRRRHDEAAVYVPPRPGPKA-GGPGVGGQVQWQON----- 203

QY 71 VMRFLOTORLPGVLDAAOARSDNCVSVLEQLPPKAAALANDRSGPPAREALAIYVF 130

DB 204 -----ASRPSDNHVVY--SLPPRPP-----SP----- 225

QY 131 GROPOPVNSLVGRLPHPSYMRDVTYERHGGPLRYHRRPLPQOE-----YLDIDOMFN- 185

DB 226 -KKPP-----PPPPAFWS-----SGSDYSDLPVLPSPGLVIGESKSTFTY 270

QY 186 RELPOA-----SGLLHNCF-YKHGRNLVTMTAPRGLOSGDRATWFGLYUNISGAF 238

DB 271 EELSRATNGFSEANLDOGGFGYHKGLPSGKREVAVKOLKAG-----SOGCE 318

QY 239 FLHHVGLLEL--VNHKALDPAKWTQKVFYQGRYDSLQLEAFEGALVNVYL--IPD 293

DB 319 REFOAEVEIISRVRH-----HLVSLIGYCMAGVOGLLVEFEVFN 358

QY 294 N-----GTGC-----SWSLKSPPVPPGAPPLQFYQGFPSVQGSRASSLWTSFGLGA 343

DB 359 NMLEFHLHGKRPTEMWSTRLKIALGSKKGLST-----LHE 394

QY 344 FSGPRIDVRFQGERLYVEISLQDALIYCGNSPAATTTRYVD-----GGFG-----M 391

DB 395 DGNPKIITHROIKASNIILDFEFKAVADFGAKIASDNTNHVSTRWGTFOYLAPRYAAS 454

QY 392 GKVTTP---LTRGVDCPYLAT-----YVDW----- 413

DB 455 GKLEKSDVFSFGVLLLELIGRRPVANNVYVDSLDVMAKPLLNASEGDEGLADS 514

QY 414 -----HFLLESGAPKRTINDAFV-----FEONGLP-----LRHHSDLY 448

DB 515 KMGNEIDREEMARVACAAACVRRSARRRPMGSIYRALBSNWSLSLDNEGRRGHSNVY 574

Tue May 20 13:18:49 2003

us-10-081-408-2.rail

Page 1

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 24.5906 Seconds
(without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095

Sequence: 1 MNQKTLVLLILAVITIFAL.....QAACAPDLPAFSGHGSN 763

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*

2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*

3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*

4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*

5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*

6: /cgn2_6/ptodata/2/1aa/backfilee1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	2.6	617	US-09-063-733A-58	Sequence 58, Appl
2	102.5	2.5	541	US-09-158-767-19	Sequence 19, Appl
3	102.5	2.5	541	US-09-158-767-20	Sequence 20, Appl
4	101	2.5	617	US-09-063-733A-16	Sequence 46, Appl
5	100.5	2.5	1253	US-08-506-296B-14	Sequence 14, Appl
6	100	2.4	617	US-09-314-242-2	Sequence 2, Appl
7	95	2.3	617	US-09-063-733A-57	Sequence 57, Appl
8	94.5	2.3	1751	US-09-136-574A-44	Sequence 44, Appl
9	94.5	2.3	15281	US-08-471-119A-2	Sequence 2, Appl
10	93.5	2.3	485	PCT-US94-03437-8	Sequence 8, Appl
11	93.5	2.3	486	US-07-737-071A-2	Sequence 2, Appl
12	93.5	2.3	486	US-08-022-096-2	Sequence 2, Appl
13	93.5	2.3	486	US-08-445-463B-8	Sequence 8, Appl
14	93.5	2.3	486	US-08-445-464C-8	Sequence 8, Appl
15	93.5	2.3	486	US-08-044-857D-8	Sequence 8, Appl
16	93	2.3	1447	US-08-540-406-19	Sequence 19, Appl
17	93	2.3	1447	US-08-656-055-19	Sequence 19, Appl
18	93	2.3	1447	US-08-954-668-19	Sequence 19, Appl
19	93	2.3	1447	US-09-268-140-5	Sequence 5, Appl
20	93	2.3	1447	US-08-918-658-19	Sequence 19, Appl
21	93	2.3	1447	PCT-US95-13233-19	Sequence 19, Appl
22	93	2.3	2972	US-09-579-181-2	Sequence 2, Appl
23	93	2.3	3118	US-08-078-311-12	Sequence 12, Appl
24	92.5	2.3	448	US-08-460-402-12	Sequence 12, Appl
25	92.5	2.3	493	US-08-078-311-7	Sequence 7, Appl
26	92.5	2.3	493	US-08-460-402-7	Sequence 7, Appl
27	92.5	2.3	493	US-08-460-402-7	Sequence 7, Appl

28	92.5	2.3	635	1	US-08-184-327A-4	Sequence 4, Appl
29	92.5	2.3	635	2	US-08-078-311-1	Sequence 1, Appl
30	92.5	2.3	635	5	US-08-460-402-1	Sequence 1, Appl
31	92.5	2.3	635	5	PCT-US95-00670-4	Sequence 4, Appl
32	90.5	2.2	1436	2	US-08-652-971-2	Sequence 2, Appl
33	90.5	2.2	1436	2	US-08-991-258A-2	Sequence 2, Appl
34	90.5	2.2	1436	2	US-08-769-399-2	Sequence 2, Appl
35	90.5	2.2	1436	3	US-08-991-653A-2	Sequence 3, Appl
36	90.5	2.2	1467	4	US-09-323-656-38	Sequence 38, Appl
37	90	2.2	617	1	US-09-063-733A-56	Sequence 56, Appl
38	89.5	2.2	623	4	US-08-332-838-2	Sequence 2, Appl
39	89.5	2.2	630	4	US-09-228-986-71	Sequence 71, Appl
40	89.5	2.2	911	4	US-08-460-269C-4	Sequence 4, Appl
41	89	2.2	688	4	US-08-973-273-26	Sequence 26, Appl
42	89	2.2	747	4	US-08-973-273-3	Sequence 3, Appl
43	88.5	2.2	379	3	US-09-028-934-36	Sequence 36, Appl
44	88.5	2.2	512	1	US-08-462-484-6	Sequence 6, Appl
45	88.5	2.2	512	1	US-08-441-147-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-09-063-733A-58
; Sequence 58, Application US/09063733A
; Patent No. 6372211
;
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; STREET: PO Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063/733A
; FILING DATE: 21-Apr-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-787-1440
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
;
; US-09-063-733A-58
;
; Query Match 2.6%, Score 106; DB 4; Length 617;
; Best local similarity 22.8%, Pred. No. 0.073;
; Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;
;
; OY 63 LSREELTAVMFLTORGLG-PGLVDAAQAR-----PSDNCVFSVEIQLPKAA 109
; DB 18 LIREVVAELNITLRLGLEVPICPPNRCGRFLGSGYSHDNLPDLYSSIK----PGCGG 73
```

```

Query Match      2.5%; Score 102.5; DB 4; Length 541;
Best Local Similarity 19.6%; Pred. No. 0.14;
Matches 109; Conservative 60; Mismatches 205; Indels 183; Gaps 22.

QY    170 PVLFOEYLDIDIMINRELPOASGLLHNCFF----YKNGRMLVMTTAPRGLOSGDRAT 225
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    40 PELVGHAEDEMHMWING-NRRAGTGYYTCIFAVPGVARNG-GLVYTICPRNLEHYLKAR 97

QY    226 -----WEGLYYNISGAGF-----LHHVGIELLVNHKAIDPARMWTOKYFY 266
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    98 PDNYEKGPFWHGCVFDLLDGDGFNSDGGDTWLAKRKTALEFTRTLRLTRAMSRWSKSI-- 155

QY    267 QGRYYSLAQL---RAQFEAGLVNVYLIPDNTOGSWSLSKVPPRPAPPLDPYQGPRF 323
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    156 HORLLPILDAADAKGAQVDOLDLLRLTFDNLCGAFC----KDPETLAAGLPENEPFA 209

QY    324 SVQGSRVASSLTWFSF-----GLGAFSGPRIFDVRFQGERLYVEISIQEALAY 372
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    210 SAFDATEATEYTLRFITFPFERLMCRKMWLG-----METTLTSSMAHV 251

QY    373 GGNSPAAMTTTRYVDGFGNGKYTPLTR-----GYDC 404
      | : | : | | | | | | | | | | | | | | | | | | | | | |
Db    252 DQYLAAYIKKKRLLELAGNGKCDPMATHDDLSREPMRKSGSYSDSLQHVALMFILAIGRT 311

QY    405 PYLATIYVDMMHFLLESQAP---KTIRDAFCVEQNOGL-----PIRRHSIDLXSHY 451

```

```

Query Match 2.5%; Score 10.02.5; DB 4; Length 541.
Best Local Similarity 19.6%; Pred. No. 0.14;
Matches 109; Conservative 60; Mismatches 205; Indels 183; Gaps 22.

QY 170 PVLFOEYLDIDIMENRELPOASGLLHNCFF---YKHGRNLVMTTTPRGLQSGDRAT 225
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 40 PELVGHAEEDMEHWING-NLRAGATGYOTCIFAVPVARGG-GLVTVTCOPRNLEHYLKAR 97
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 226 -----WFLYYNISGAGFF-----LHHVGLLELLVNHKALDPAWTTIOKVR 266
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 98 FDNTEKGFPMHGVFRLDLDDGIFNSDGDWTMLAQRTALAEFTTTLRTLAMSRYVNSKI-- 155
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 267 OGRYVDSLALQ---EAFQAGLVNVVLLPDNNGSGWSLKSVPKPCAPPLQFYPCGPRF 323
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 156 HGRLLPILDAKAKKQAVDLDLLRLRTFDNLCGLAFG-----KDPETLAGLPEPNEEA 209
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 324 SVQGSRVASSLWTFSP-----GLGAFSGPRIFDVRQGERLTVEISIQEALATY 372
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 210 SADFDATELTLLRFLFFPEFLMRCKKMWLGIG-----METTLTSSMAHY 251
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 373 GGNSPALMTTRYVDGFGFGKTYTTPLTR-----GVDC 404
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 252 DQYLAIVIKRKRLTELAGNGKCDTATHDLLSRFKRKSYSDESLOHVALNFIILAGBPT 311
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 405 PLYLATYVDMHFLFLEQAP---KITRDAFCVPEQNGL-----PLRRHSDLYSRY 451
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 312 SSVA--LSMFEPVLVSTHPVERKIVRELSCTVLAASGAHPALMLTAEPFTFEELDLYVL 369
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

OY 452 FGLAETLVVRSM-----STLNTVDYVMDYVHPSG-AIEIRFYATGYISSAFLEGATGK 506
DB 370 KAASETILRLPYVEDSKHVVADYLDPGTFVPAGSSVTYSYAGARRKG----- 420
OY 507 YGNVSEHTLGTVHTSHAFKVDLDVAGLENNVMAEDWVFVMAVPMSPHQLOQLQVTR 566
DB 421 -----VMGED-----CIEFRE-----RWLSADG 439
OY 567 KLEMEQAPLVGATERYLYLASNSKMGHPRGYRIQMLSFAGEPLPQNSSMARGFS 626
DB 440 TKFEQHDSDYKFAVFNAGRVCL-----GKDLAY-LQMKNIAGSVLLRH----- 481
OY 627 MERYOLAVTORKEEPS 643
DB 482 --RLTVAPOGHRVEOKMS 496

RESULT 4

US-09-063-733A-46
Sequence 46, Application US/09063733A
Patent No. 6372211

GENERAL INFORMATION:
APPLICANT: Isaac, Barbara G.
APPLICANT: Greenplate, John T.
APPLICANT: Purcell, John P.
APPLICANT: Romano, Charles P.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold White & Durkee
STREET: PO Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/063,733A
FILING DATE: 21-APR-1998
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Paterson, Melinda L.
REGISTRATION NUMBER: 33,062
REFERENCE/DOCKET NUMBER: MOBT:022
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-787-1440

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:
LENGTH: 617 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
US-09-063-733A-46

Query Match 2.5%; Score 101; DB 4; Length 617;
Best Local Similarity 23.3%; Pred. No. 0.25;

Matches 81; Conservative 39; Mismatches 138; Indels 90; Gaps 19;

OY 63 LSRHELTAVMRFLTORIG--PGLYDAQAAR-----PSDNCVFVSEYLOLPPKAA 109
DB 18 LIREKXAKELNITERLGEVGPIPPNEGFLGSGYSHDLPSDPLYSST-----KPL 71
OY 110 LAHLIDGSP--PPAREALAVFFGROPQPNVSELVG-----PLPHPSIMRDVYTERRG 161
DB 72 LKEAPRAEBELPPRK--VCIVGAG-----VSLGIYAMILDKIDNLTIDIFESSRTG 123

OY 162 GPLYHRRPVLVEQYLDIDOMIFNELPOASGLHHCCEYKRGKGNLVMTAPRLQSG 221
DB 124 GRLTYHHTDAKHDYIDGAMRY-----PDIPMKATFNLFKRTGAPLI-----KYLDG 173
OY 222 DRATWFGLYYNISGAGFELHVGLELVN-----HKAIDPARMTQKYFYQGRYDS 273
DB 174 ENTIP--QLYNN-----HFAKGVSDPYMWVANGVTPDDVDVSVEKILQOAF--GYKKEK 225
OY 274 LAOLEAOFELAVNVLLIPDNGTGSWSLSKSPVPPGAPPLQFYQGPFSYQSRVASS 333
DB 226 LAE--DEDKGFDELMLYDDMTT--REYLKRGPPGAEAPKXDF-----FAIQMETON- 273
OY 334 LWTFSGLGAFSGPRIFDVRFQGERLVEISLQ-----EALAYGNS 376
DB 274 -----TGTNLFDOAF--SESYIDSFDDNPNKPMYCIEGGTS 309

RESULT 5

US-08-506-296B-14
Sequence 14, Application US/08506296B
Patent No. 6313265

GENERAL INFORMATION:
APPLICANT: Phillips, Greg
APPLICANT: Cunningham, Bruce A.
APPLICANT: Crossin, Kathryn L.
TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute
STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
CITY: La Jolla
STATE: California
COUNTRY: U.S.
ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506,296B
FILING DATE: 24-JUL-1995
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 1253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-506-296B-14

Query Match 2.5%; Score 100.5; DB 4; Length 1253;
Best Local Similarity 18.5%; Pred. No. 0.95;

Matches 160; Conservative 89; Mismatches 269; Indels 349; Gaps 40;

OY 23 VLVYRGCGG-----GEPSQLPHCPVSPSAQP-WTHPGSQQLFA----- 61
DB 299 LKVGEDDEGGEYRCLAEINSIGSARHAYVYVEAPYMLHRPOSHLYGPGETARLDQYOG 358
OY 62 -----DISREELAVMRFLTORIGPGLYDAQAARSPNCFVSVE----- 100
DB 359 RPOEYTWKRIQIVVEELAKDKQYRIQR--GALLISNVOQSPSTWYTOCEARRHGLLLA 415
OY 101 -----LOLPPKAAALAHLDGSPPPAREALAVFFGROPQPNVSELVGVLPHPSYMD 154

```

Db 416 NAVIYVQLPAKILITADNOTYMAVOGSTAYLLCKAFG-APVPSV----- 458
QY 155 VYVERHGGLPYHRRPVLFQBYLDIDOMIF---NRELPOASGLLHCCFYKHKRNLYTM 211
Db 459 -----QWLEDEGTTLQDERFFPYANG----- 480
QY 212 TTAAPRGLOSDBRATWGLYYNIGAGFELHVGLELLVNHKALDPAWTKQVYQGRY 271
Db 481 TLGIRLDQANDTGRYFCLANDON-----NNTIMANKVKDATOT----- 521
QY 272 DSLAQLAEQAGLVNVVLIIPDNGTGSWSLSKSPVPPAPPLQFYQPPRFVSQ---GSR 329
Db 522 -----QGRSTIEKKGR 534
QY 330 VASSLMTFFSGGLAFSGPRIFD-----VRPQGE-RLVYEISLOELALATYGCNSPAAWTT 382
Db 535 V-----TFTCOASFDPSLQPSITWRGDRLOELGSDKXFIEDGRL-VIHSL 581
QY 383 RYVDGFGMGKYYTPLRGRVDCPYLATYVDMHFLLESQA-----PKTRDAFCV 431
Db 582 DYSD---QGNYSQVASTEID-----VESRAQLLVYSGPVPRLVLSDLHL 625
QY 432 FEQNO---GLPLRRHSHDYSHYFGGLAETVLVVRSMSTLLNDYVDTVFHPSGAIEI 487
Db 626 LTQOVHVSMSPADHNAPIEK-----YDIEFEDKEMAPE 660
QY 488 RFIYATG-----YISSAFLFGATGKYG---NOVSEHLGTYHT-HSAHF 527
Db 661 KWSYLGKVPNGQSTTKLSPIYHYTRVTAINKYGRQEPVSE---TYVTEPAPEK 716
QY 528 VDLVAGLENNVMAEDVVFVMAVWSP-----EHQLO-RLQVTKLLEMEQA 575
Db 717 NPVYVKG-----EGNETTMVITWKPLRMMDNNAPOVYRVQRRQGTG--PMQOI 767
QY 576 A---FLVGSATPRLYLASNHSNKGHPRGYRIQMLSFAG-PLPONSMAFGSWERYQ 631
Db 768 VSDPFLVSNSTFV-----PYEIKVOAVNSQKGPQ---YVIGSGEDYP 812
QY 632 LAVQKREEESSSVFNQNDPAPTYVDFSPINNETIAGDLYAMVYTAG-----FLH 684
Db 813 QATLELEIETILNLSAVLYK--WRP-VDLAOKGH--LRGINVYWRREGSKRKHSHIH 867
QY 685 IPHAEDIPNTVTVNGVGFPLRPYNFDEDEDSFYASDSIYRGDODA-----G 732
Db 868 KDHYVVPANTSV---ILSGLRPYSYILEVOAENG-----RGSPPASEFTFSPTEGVPG 919
QY 733 ACEVNPPLACLPQAAACAPDLPFESHG 759
Db 920 HPEALHLECCQSGNTSILLRMQPLSHNG 946

RESULT 6
US-09-314-242-2
: Sequence 2, Application US/09314242A
: Patent No. 6248575
: GENERAL INFORMATION:
: APPLICANT: Elizabeth J. Gollightly
: TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
: FILE OF INVENTION: Having L-Amino Acid Oxidase Activity
: CURRENT APPLICATION NUMBER: US/09/314,242A
: EARLIER FILING DATE: 1999-05-18
: EARLIER APPLICATION NUMBER: 09/080,428
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 617
: TYPE: PRT
: ORGANISM: Trichoderma harzianum
US-09-314-242-2
Query Match 2.4%; Score 100; DB 4; Length 617;

```

```

Best Local Similarity 23.9%; Pred. No. 0.31;
Matches 83; Conservative 36; Mismatches 139; Indels 90; Gaps 18;

QY 63 LSREELTAVMKRFLLQRLG--PGLVDAQAAR-----PSDNCVSVELQPPRAA 109
Db 18 LIREKVAKEINILTERGEVPIPPNEGREFLAGGYSHDMLPSDPLYSI-----KPAL 71
QY 110 LAHLDGRSP--PPAREALATVFGROPQPNVSELYVG-----PLPSPYMRDVTYVRHG 161
Db 72 LKEAPRAEEELPPRK--VCIVGAG-----VSGLYIMIIDDLKIPMLTYDIESSSRTG 123
QY 162 GPLPYHRRPVLFQBYLDIDOMIFNRELPOASGLLHCCFYKHKRNLYTMTAPRGLOSG 221
Db 124 GRLYTHFTDAKHXYIIGAMRY--PDIPSMK-----RFFNLFKTKKPLIYYL 171
QY 222 DRATWFLGLYINISAGFELHHVGLLELVN-----HKALDPAWTKQVYQGRYSD 273
Db 172 DGEWTPQLYNN---HFFAKGVSDPYWVSANGSTVPDDVDVSXGKELQOAF--GYK 225
QY 274 LAQLEAPQAGLVNVVLIIPDNGTGSWSLSKSPVPPAPPLQFYQPPRFVSQSRVASS 333
Db 226 LAE--DFDKGFDELMLVDWTT--REYLKRGKGPGEAPKYDF-----FAIQMETON- 273
QY 334 LMTFSFGIGAFSGPRIFDVRFQGERLYEISLQ-----EALATYGCNS 376
Db 274 -----TGYNLFDAQF-SESVIDSFPDNPTRKPMWYCIEGTS 309

RESULT 7
US-09-063-733A-57
: Sequence 57, Application US/09063733A
: Patent No. 6372211
: GENERAL INFORMATION:
: APPLICANT: Isaac, Barbara G.
: APPLICANT: Greenplate, John T.
: APPLICANT: Purcell, John P.
: APPLICANT: Romano, Charles P.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold White & Durkee
: STREET: PO Box 4433
: CITY: Houston
: STATE: TX
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/063,733A
: FILING DATE: 21-APR-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Patterson, Melinda L.
: REGISTRATION NUMBER: 33,062
: REFERENCE/DOCKET NUMBER: MOBT-022
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 713-787-1400
: TELEFAX: 713-787-1440
: INFORMATION FOR SEQ ID NO: 57:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 617 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: linear
US-09-063-733A-57
Query Match 2.3%; Score 95; DB 4; Length 617;
Best Local Similarity 22.5%; Pred. No. 1;

```

Matches 78; Conservative 40; Mismatches 142; Indels 86; Gaps 18;

QY 63 LSSEELTAVNRFLTLRLG--PGLVDAQAAR-----PSDNCVFSVELDLPKAA 109
 Db 18 LIREKAKELNITLERLGEVPIPPNEGRLGGYSHDLPSDLYSSI----GGSSGG 73

QY 110 LAHLIDGSPPRAREALAIYFPGROPQPNVSELYVG-----PLPHPSYMDVYTERGCP 163
 Db 74 XXXXXGGGPPRR--VCIVAG-----VSGLYTAMLLDCLKIPNLTYDIFESSRRGR 125

QY 164 LPHRRPRLVFOEYLDIDOMIFNRELPOASGLLHHCFCYKRGRLVTMTAPRGLQSDR 223
 Db 126 LYHHHFTDAKHDDYIDCAMY-----PDIPSKRTFNLFKRTGMLI-----KYLDGEN 175

QY 224 ATWEGLYNISGAGFLHVGLELVN-----HKALDPAWMTIQKVFYQGRYDLSA 275
 Db 176 TP--QLYNN-----HFEKAGVSDPMVSVANGTVPDVDVSVGEKLQAAF--GYKKERLA 227

QY 276 QLEAOPAGLVNVLIPDNGTSGSWLSKSPVPGAPPLQFYPGCRPSVSGSVASSLW 335
 Db 228 E---DEFGKDELVDMDMT--REYIKRGKPGKEAPKIDP-----FALQMETON--- 273

QY 336 TFEGLGAFSGPRIFDVRFOGERLYVEISLQ-----EALAIYGGNS 376
 Db 274 -----TGTNLFDOAF--SESVYIDSPDFDNPTRKPEMYCIEGCTS 309

RESULT 8
 US-09-136-574A-44
 ; Sequence 44, Application US/09136574A
 ; Patent No. 6294366
 ; GENERAL INFORMATION:
 ; APPLICANT: Farrington, Graham K.
 ; Anderson, Paige
 ; Gibbs, Moreland
 ; Bergquist, Peter
 ; Daniels, Roy
 ; Morgan, Hugh W.
 ; Williams, Diane P.
 ; TITLE OF INVENTION: Compositions and Methods for
 ; Treating Cellulose Containing Fabrics Using Truncated
 ; Cellulase Enzyme Compositions
 ; NUMBER OF SEQUENCES: 49
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Howson and Howson
 ; STREET: Spring House Corporate Center, P.O. Box 457
 ; CITY: Spring House
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19477
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/136,574A
 ; FILING DATE: 19-Aug-1998
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/932,571
 ; FILING DATE: September 19, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bak, Mary E.
 ; REGISTRATION NUMBER: 31,215
 ; REFERENCE/DOCKET NUMBER: 1997US001/CIP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-540-9200
 ; TELEFAX: 215-540-5818
 ; TELEX: <Unknown>
 ; INFORMATION FOR SEQ ID NO: 44:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1751 amino acids
 ; TYPE: amino acid

STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 44:
 US-09-136-574A-44

Query Match 2.38; Score 94.5; DB 4; Length 1751;
 Best Local Similarity 19.08; Pred. No. 7.3; Indels 225; Gaps 35;
 Matches 139; Conservative 83; Mismatches 284;

QY 29 GGDGERSQPLPHCSVSPSAQPM--THPGSQLADREELTAVNRFLTLRLGRLVD 85
 Db 578 GGEVEHKKQAFKISV--PGGYPMPTNDPSYKGLTSOLEKKYIA----- 621

QY 86 AAOARPSDNCVFSVELDLPKAAALHLDRGSPPRAREALAIYFPGROPQPNVSELYVG 145
 Db 622 ---AYDNNNLWGLE-----PGAASTPAPTSTPPT-----PTPTVTATPT-P 663

QY 146 LPHPS-----YARDVTERHGGRLPHRRPVLFOEYLDIDOM-----IFN 185
 Db 664 TPPTPTGSPGTGSGVAVLKNNETASSTGIRPWFKIVNGSSVDSRKIKRYWTV 723

QY 186 RELPOASGLLHHCFCYKRGRLVTMTAPRGLQSDRATWGLYUNISGAGFLHVG 245
 Db 724 GDRQSA-----VCDMAIGASNTFNFVK--LSSG-----VSGADYLL-EGVF 764

QY 246 ELVNHKALDPAWMT--IQKVFYQGRY-----YSLAOLEAOFENG-----LVN 291
 Db 765 SSGAQO--LQPKDGTGDIQVRFNKNDSNYNQADDSWLSQMTYGENAKVTLV 822

QY 292 ---PDNGTSGSWLSKSVP-----PGAP-P-PLQFYPGCRPSVSGSVASSLW 339
 Db 823 WGPBPGATPAPTSTAPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 882

QY 340 GLGAFS-----GPRIFDVRFOGERLYVEI--SLOEALAIYG 373
 Db 883 SYGALKWYANGNLSSPTNVLNPKIKIENVTGTAVIDLSRVKVRWTTIDGSAIOSVSAS 942

QY 374 GNSPAAWTRRYVDCGFGNGKATTPLTRGVDCPYLATIYDWHF-----LLESQAPK 426
 Db 943 SINPAYIDVKFKLGANAG-----GAD-----YVEIEFGSGAGVLAQOSTKEIRL 989

QY 427 ---DAFCVEONOGPLRHHSDLSYHFGLAETVLVVRSMSTLANNDYMDVFRHSG 483
 Db 990 SIQKSGSYNOSNDISVANASYSIENEKVYIDVL-----YMGK--EFGK 1034

QY 484 AIEIR-FYATGISAEFLGATGKYGNOSVSHLTGTVTHSAHFVLDLVAGLENNWAAE 542
 Db 1035 NAOIKWYANGNLGS-----MTNVLNPKIKIENVTG-----AVDLSRVKVRWTTID 1082

QY 543 DMVEFPAWVPSPEHOLQLOVTRKLLMEQAAFLVGSATPRVL--YLASNHSKMGHP 600
 Db 1083 GEA-----TQSVSVTSINPAYIDVVFVRLGANAGAD 1115

QY 601 RGYRIQMLSFAGEPLPQNSMARGPSMERCYLATQOREEPPSSSVFNDDNPAPTYDF 660
 Db 1116 YVEIEGFKSGAG-----VLAAGOSTKEIRLSTOK-----GSGSYNOSNDIS-VASA 1160

QY 661 SDFINNETIAG 671
 Db 1161 TGYIENEKVYG 1171

RESULT 9
 US-08-471-119A-2
 ; Sequence 2, Application US/08471119A
 ; Patent No. 5827706
 ; GENERAL INFORMATION:
 ; APPLICANT: Leitner, Ernst
 ; APPLICANT: Schneider, Elisabeth
 ; APPLICANT: Schoeenderfer, Kurt
 ; APPLICANT: Weber, Gerhard
 ; TITLE OF INVENTION: Cyclosporin Synthetase

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5827706artlis Corporation
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,119A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kasenoff, Melvyn
REGISTRATION NUMBER: 26,389
REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 503 8474
TELEFAX: 201 503 8807
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 15281 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Tolypocladium niveum
STRAIN: ATCC 34921
US-08-471-119A-2

Query Match 2.3%; Score 94.5; DB 2; Length 15281;
Best Local Similarity 18.1%; Pred. No. 3.2e+02;
Matches 154; Conservative 110; Mismatches 307; Indels 279; Gaps 41;

QY 45 SPSPAPWTHPGOSOLFADLSREELTAVNRFLTQRL-GPGLYDAQAQPSDNCVFSVEIQL 103
DB 12682 NPDAIAVDTSTKLYLADQSDHVAWLSKQKLPASIVVLA PRSSETIVACIGI- 12739
QY 104 PRKAAALAHLDGRSPRRAREALATVFEGROPQPNVSELVGPVLPHPSTYKRDVTVERHGGP 163
DB 12740 --LAAALAYLWDSNVPRARQAIL---SEIPGEKFFVLLGA-----GVP 12778
QY 164 LPYHRRPVLPQEVLDIDQMFNRELPAQASGLLHCCFYKHKGRNLVTMTTARGLQSGDR 223
DB 12779 IPDKNT-----ADV-RMVFISD-----IVASKTDKSTSPGTR 12809
QY 224 ATWEGLYVNI--SGA-----GFFLHNGVLELL-----VNHKALDPARMT 260
DB 12810 PSASLAVVITFGSGTGPRKGVMEHGRVISLQKNASRIPQSLMAHVNLAFLASVWE 12869
QY 261 IQKVFYQG-----RYT---DSLQALEQFEAGLVNVVILIPDNGTGGSKLSKVPVPGP- 310
DB 12870 IFTLLNGTGLFCISYFVLVDSKA--LSAASFSDHRINITLLPPA-----LLKQCLADAPS 12922
QY 311 -----APPLQFYPOGPRPSVQG-----SRVASSLM-----TFSEGL-L 341
DB 12923 VLSSLESLXIGDRLDQADATKVKLVKQKAYNAVGPENSNWSTIYLTETHEFANGVPI 12982
QY 342 GAFSGPR---IFDVRFQGERLVEISLQELALATVGNSPAAMTTRYVDGFGMKKTYTP 397
DB 12983 GTSLEPKRSKAYIMD---ODOQLV-----PACVMEGLVAVDGG----- 13016
QY 398 LTRGVDCYUATLYVDMHRLLESQAPKTRDAFCVEQNOGRLPRRHHSLYSHYFGGLAE 457
DB 13017 LARGTDPSSLNTGRFTHITIDCKQVOAYR-----TGDRVRYRPRDYOIEFFGRILDQ 13067

QY 458 TLVYVRSMSTLLNDYVMDVTFHPGSAIEIRFYATGYISSAFLEGGATGKGNQV----- 511
DB 13068 QIKIRGHR-----IEPA-EVEQALLSDSSINDVAVVSAQNKEGLEMGYITTT 13113
QY 512 -----SEHTLGVTHVTHSAHFK-----VLDVAGLE--MMVAEDMYFVPMAYPM 553
DB 13114 QAAQSVDEKEASNNQVQEWFAHFDSTAYANIGIDBDALGQDFLSTWSDSLIPR----- 13169
QY 554 SPEHOLRLQYT-RKLEMEQDAFL-VGSA TPRLYTLASHNSKNGKHPRGYRIOMLSFA 611
DB 13170 --EEMQEWLNDMTMRSLDNOFPKVLKIGTGMVLF-----NLGKVEEG---LQSYA 13216
QY 612 GEPILQNSMARGFEMERYQLAVTQRKEEPPSSSVFNQNDPMAPTVPFSPFINNETTA- 670
DB 13217 G--LEPSRSVT---AW-----VNKAITFPPLASARVHVGTAEDISSIDGLRSDLVVI 13265
QY 671 -----GKDLVAVWTAAGFLHPHAEIDIPMTVTGNGVGFF--LRPY-----NF----- 710
DB 13266 NSVAGYFPSPRELAETLANLRLPGVKRI-----FFGDMRTYATKNDFLVARAV 13314
QY 711 -----FDEDPSTYSADSIYFRGDDAGACEVNPPLACLPQAAACAPDL 752
DB 13315 HTLGSNAKAWYRQGVAKLEDEDELLVDPAFETSLSDQFPDEIKHVELPRRMAATNEL 13374
QY 753 PAFSHGFSH 762
DB 13375 SYRYAAVTH 13384

RESULT 10
PCT-US94-03437-8
Sequence 8, Application PC/TUS9403437
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HOMONEOUS IMMUNOASSAYS USING MUTANT
TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASES
NUMBER OF SEQUENCES: 124
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/03437
FILING DATE:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Leuconostoc dextranicum
STRAIN: ATCC 19255
PCT-US94-03437-8

Query Match 2.3%; Score 93.5; DB 5; Length 485;
Best Local Similarity 20.8%; Pred. No. 0.99; Indels 97; Gaps 20;
Matches 81; Conservative 61; Mismatches 151;

QY 406 YLATYVDWHF-----LLESQAPKTRDAFCVEQNOGRLPRRHHSLYSHYFGGLA 456
DB 30 YKKGYLEHFAIVGTAROO LSDDEFKQLVRSIKDFTEDQA-----QAEAFIAHFSYRA 83
QY 457 ETVLVVRSMSTL-----LNDYVMDVTFHPGSAIEIRFYAT--GYISSAFLEGGATGK 506
DB 84 HDVTPDAASYGILKSAIEEATKFDIDGNRIFYMS--VAPRFEGTAKYLSKSGLLAETG- 140


```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,464C
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/044,857
FILING DATE: 08-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Ruzsala, Lois K.
REGISTRATION NUMBER: 39,074
REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 267-5364
TELEFAX: (847) 267-5376
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYDROTHERICAL: NO
ANTI-SENSE: NO
US-08-445-464C-8

Query Match      2.3% Score 93.5; DB 3; Length 486;
Best Local Similarity 20.8%; Pred. No. 1;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

QY 406 YLATYVDHMF-----LLESQAPKTIRDAFCVEONQGLPLRRHSDLSHYFGCLA 456
DB 31 YKKGIIQEHFAIYGTARQOLSDDEFKQVYDSIKDFTEDA-----QAEAFIAHFSYRA 84
QY 457 ETVLVVRSNSTL-----LNTDYWDVTFHPSGAIEIRFYAT--GYISSAFLFGATGK 506
DB 85 HDVTDAAASYGILKSAIEEATKFDIDGNRIFYMS--VAPRFEGTIAKYLKSEGLAETG- 141
QY 507 YGNQVSEHILGYTHSAHFHKVDLDVAGLENWV-----AEDMVFPMAV----- 551
DB 142 YNRIMIEKPFGTSYATAEELQSDLENAFDDOLFRIIDHYLGKEMVONIALRFGNPIFDA 201
QY 552 PMSPEHQLORLOYT-RKLEMEBOAFLVGSATPRYLYLASNHSNKGHPRGYRIOMLSF 610
DB 202 AMNKDY-IRKNOVTLAEVLGVEERAGY--DTTCALLDMQNH-----MQLVGM 248
QY 611 AGEPLPQ--NSSMARGFSWERYOLAVTORKEEPESSSVFNQNDPMAPYVDESDF----- 664
DB 249 LAMEKPEFNDKDIRAKNAAFN-ALKIYNEEVNKYFVRAOYGA-GDTADYKPYLEAD 306
QY 665 -----NNETIAGK---DLVAVYTAGFLHPHADIPNTYVGN-----GYGFELR 706
DB 307 VPADSKNNTFIAGELQDFLPRW-----EGVFPYVSGRLAAKOTRVDIVFKAG 355
QY 707 PYNFDEDPFSYSADSIYFRGDODAGACEV 736
DB 356 TPNFGSEQAQESVLSIT--DPKGAIEL 382

RESULT 15
US-08-044-857D-8
Sequence 8, Application US/08044857D
Patent No. 6455288
GENERAL INFORMATION:
APPLICANT: Jakobovits, Edward B.
Silen, Joy L.
Levy, Mark J.
Goodman, Thomas C.
Becker, Martin
```

```
Caldwell, Robert M.
Bott, Richard R.
Barnett, Christopher C.
TITLE OF INVENTION: Homogeneous Immunoassays Using Mutant
Glucose-6-Phosphate Dehydrogenases
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dade Behring Inc.
STREET: 1717 Deerfield Road
CITY: Deerfield
STATE: Illinois
COUNTRY: USA
ZIP: 60015

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,857D
FILING DATE: 08-APR-1993
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ruzsala, Lois K.
REGISTRATION NUMBER: 39,074
REFERENCE/DOCKET NUMBER: BEH-7261
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 267-5364
TELEFAX: (847) 267-5376
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYDROTHERICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Leuconostoc dextranctum
STRAIN: ATCC 19255
SEQUENCE DESCRIPTION: SEQ ID NO: 8;
US-08-044-857D-8

Query Match      2.3% Score 93.5; DB 4; Length 486;
Best Local Similarity 20.8%; Pred. No. 1;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

QY 406 YLATYVDHMF-----LLESQAPKTIRDAFCVEONQGLPLRRHSDLSHYFGCLA 456
DB 31 YKKGIIQEHFAIYGTARQOLSDDEFKQVYDSIKDFTEDA-----QAEAFIAHFSYRA 84
QY 457 ETVLVVRSNSTL-----LNTDYWDVTFHPSGAIEIRFYAT--GYISSAFLFGATGK 506
DB 85 HDVTDAAASYGILKSAIEEATKFDIDGNRIFYMS--VAPRFEGTIAKYLKSEGLAETG- 141
QY 507 YGNQVSEHILGYTHSAHFHKVDLDVAGLENWV-----AEDMVFPMAV----- 551
DB 142 YNRIMIEKPFGTSYATAEELQSDLENAFDDOLFRIIDHYLGKEMVONIALRFGNPIFDA 201
QY 552 PMSPEHQLORLOYT-RKLEMEBOAFLVGSATPRYLYLASNHSNKGHPRGYRIOMLSF 610
DB 202 AMNKDY-IRKNOVTLAEVLGVEERAGY--DTTCALLDMQNH-----MQLVGM 248
QY 611 AGEPLPQ--NSSMARGFSWERYOLAVTORKEEPESSSVFNQNDPMAPYVDESDF----- 664
DB 249 LAMEKPEFNDKDIRAKNAAFN-ALKIYNEEVNKYFVRAOYGA-GDTADYKPYLEAD 306
QY 665 -----NNETIAGK---DLVAVYTAGFLHPHADIPNTYVGN-----GYGFELR 706
DB 307 VPADSKNNTFIAGELQDFLPRW-----EGVFPYVSGRLAAKOTRVDIVFKAG 355
QY 707 PYNFDEDPFSYSADSIYFRGDODAGACEV 736
```

Db 356 TFNFGSEGEAQESVLSIIT---DPKGAIEL 382

Search completed: May 20, 2003, 11:40:55
Job time : 36.5906 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 121.567 Seconds

(without alignments)
622.443 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095
Sequence: 1 MNOKTILVLLAVITIFAL.....QMAACAPDLPAFSGHGSFN 763

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 9917265 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PTI_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PTI_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/METHOD_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4095	100.0	763	9	US-10-081-408-2
2	3967	96.9	998	9	US-10-081-408-20
3	2536.5	61.9	729	10	US-09-919-497-51
4	106	2.6	617	9	US-10-005-530-56
5	104	2.5	980	9	US-09-908-193-22
6	103	2.5	2224	9	US-10-115-563-14
7	102	2.5	1194	9	US-09-738-626-5832
8	101	2.5	617	9	US-10-005-530-46
9	99	2.4	944	9	US-10-213-990-27
10	95.5	2.3	581	9	US-09-738-626-4914
11	95	2.3	617	9	US-10-005-530-57
12	93.5	2.3	878	10	US-09-912-020-364
13	93	2.3	480	10	US-09-734-569-172
14	93	2.3	601	10	US-09-925-301-844
15	93	2.3	1447	1	US-08-954-301-19
16	93	2.3	1447	10	US-09-898-533-5
17	92.5	2.3	635	9	US-10-099-895-33
18	92	2.2	904	9	US-09-712-363-247
19	91.5	2.2	994	9	US-10-004-551-16

20	91	2.2	390	9	US-09-738-626-6558	Sequence 6558, Ap
21	91	2.2	1247	9	US-09-908-193-2	Sequence 2, Appl
22	91	2.2	2507	9	US-09-819-104A-2	Sequence 2, Appl
23	90.5	2.2	388	9	US-09-361-630-1	Sequence 1, Appl
24	90	2.2	589	10	US-09-740-041-2	Sequence 2, Appl
25	90	2.2	617	9	US-10-005-530-56	Sequence 56, Appl
26	89.5	2.2	306	9	US-10-017-910-7	Sequence 71, Appl
27	89.5	2.2	630	9	US-10-101-464A-71	Sequence 4, Appl
28	89.5	2.2	911	9	US-09-855-754-4	Sequence 2, Appl
29	89.5	2.2	911	9	US-10-227-353-4	Sequence 4, Appl
30	89.5	2.2	953	10	US-09-845-583-4	Sequence 2, Appl
31	89.5	2.2	3695	9	US-10-037-182-2	Sequence 1007, Ap
32	89	2.2	365	9	US-09-925-299-1007	Sequence 1007, Ap
33	89	2.2	365	10	US-09-925-299-1007	Sequence 34, Appl
34	89	2.2	1430	9	US-10-087-993-34	Sequence 1519, Ap
35	88.5	2.2	583	9	US-09-942-185-2	Sequence 574, App
36	88.5	2.2	616	10	US-09-925-300-1519	Sequence 574, App
37	88.5	2.2	882	9	US-10-174-580-574	Sequence 574, App
38	88.5	2.2	882	9	US-10-175-758-574	Sequence 574, App
39	88.5	2.2	882	9	US-10-175-737-574	Sequence 574, App
40	88.5	2.2	882	9	US-10-173-706-574	Sequence 574, App
41	88.5	2.2	882	9	US-10-175-738-574	Sequence 574, App
42	88.5	2.2	882	9	US-10-175-752-574	Sequence 574, App
43	88.5	2.2	882	9	US-10-176-482-574	Sequence 574, App
44	88.5	2.2	882	9	US-10-176-757-574	Sequence 574, App
45	88.5	2.2	882	9	US-10-176-913-574	Sequence 574, App

ALIGNMENTS

RESULT 1
US-10-081-408-2
Sequence 2, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abirams, n , Lars
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT FILING DATE: 2002-02-21
CURRENT APPLICATION NUMBER: US/10/081,408
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 763
TYPE: PRT
ORGANISM: Homo sapiens
US-10-081-408-2

Query Match	100.0%	Score 4095;	DB 9;	Length 763;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 763;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	MNOKTILVLLAVITIFALVGVLRGGDGEPSQLPHCSVSFSAQPTNPGOSQLF	60	
DB	1	MNOKTILVLLAVITIFALVGVLRGGDGEPSQLPHCSVSFSAQPTNPGOSQLF	60	
QY	61	ADLSREELTAVVRFLLTORGLVDAQAQRPDNCVFSELDLPKAAALAILDGSPPP	120	
DB	61	ADLSREELTAVVRFLLTORGLVDAQAQRPDNCVFSELDLPKAAALAILDGSPPP	120	
QY	121	AREALAIYFFGQPOPNSVGVPLPHPSYRDYTVERRHGPPLPHRRPVLFQEYLDID	180	
DB	121	AREALAIYFFGQPOPNSVGVPLPHPSYRDYTVERRHGPPLPHRRPVLFQEYLDID	180	
QY	181	QMIFNRELPOASGLLHHCFCYKRRGRNLVTMTARGLDGSRAVFWGLYVNIISAGFFL	240	
DB	181	QMIFNRELPOASGLLHHCFCYKRRGRNLVTMTARGLDGSRAVFWGLYVNIISAGFFL	240	

Qy	241	HHLELLELVNKKALDPA	RMTHQKFCYGRGYDLSAOL	EAOFEGLVNVVLI	IPNNGCGSN	3000	
Db	241	HHVGLELLVNKKALDPA	RMTHIQKFTYGRYDLSAOLEAO	FEGLVNVVLI	IPNNGTGSN	3000	
Qy	301	SLKSPVPBPAPLQAF	POGFRFSVQSSRAVSS	LMTFSFQJGAFSG	GRIDVRFQGERLV	3600	
Db	301	SLKSPVPBPAPLQAF	POGFRFSVQSSRAVSS	LMTFSFQJGAFSG	GRIDVRFQGERLV	3600	
Qy	361	YEISLOELALAYGNS	SPAAMTRRVYDGGFEMGK	KTPTLTFTGVC	DPYLATYVDMHFLLESQ	4200	
Db	361	YEISLOELALAYGNS	SPAAMTRRVYDGGFEMGK	KTPTLTFTGVC	DPYLATYVDMHFLLESQ	4200	
Qy	421	APKTRDAFCVFEON	NOCLPLRRHNSDLSYHF	FGLAETLVVBS	MSSTLVNRYDVMDFVFH	4800	
Db	421	APKTRDAFCVFEON	NOCLPLRRHNSDLSYHF	FGLAETLVVBS	MSSTLVNRYDVMDFVFH	4800	
Qy	481	PSGAEIRFVYATG	YISSAFLPGAGKXGNO	VSSEHTTCTVT	THSHAFVYDLDVAGLEMMVM	5400	
Db	481	PSGAEIRFVYATG	YISSAFLPGAGKXGNO	VSSEHTTCTVT	THSHAFVYDLDVAGLEMMVM	5400	
Qy	541	AEDVAFVMAVPMS	PEHOLQVTRXKLL	EMEQAAVLVS	ATPRYLYLASNSNSNMKGHP	6000	
Db	541	AEDVAFVMAVPMS	PEHOLQVTRXKLL	EMEQAAVLVS	ATPRYLYLASNSNSNMKGHP	6000	
Qy	601	RGYRIOMLSFAGE	PLPONSSARBSF	SEROLANTOKKEE	PSSSVFNONDPAAPVYDF	6600	
Db	601	RGYRIOMLSFAGE	PLPONSSARBSF	SEROLANTOKKEE	PSSSVFNONDPAAPVYDF	6600	
Qy	661	SDFINNETIACK	DLVAMVYAGFLIIP	AAEDIPNTVY	ONGVGFELR	RYNFEEDEDPSTYS	7200
Db	661	SDFINNETIACK	DLVAMVYAGFLIIP	AAEDIPNTVY	ONGVGFELR	RYNFEEDEDPSTYS	7200
Qy	721	DSIYFRGDDQD	GACEVNPPLAC	LPAAACAPDL	PAFSGGFSHN	763	
Db	721	DSIYFRGDDQD	GACEVNPPLAC	LPAAACAPDL	PAFSGGFSHN	763	

```

RESULT 20: 408-20
US-10-081-408-20
Sequence 20, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abrahams, n , Lars
APPLICANT: Nilsson, Jark
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: SE 010625-3
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: US 60/272,247
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 20
LENGTH: 998
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant construct
US-10-081-408-20

```

Query Match	Similarity	Score	DB	Length
Best Local	99.9%	Pred. No. 0		
Matches	735	Conservative	1	Mismatches 0; Indels 0; Gaps 0;
Qy	28	RGGDGGEPSQLHPCHSVSPSAQPTWTHPCGOSQLFADLSREELTAVMKFLQRLGPGTVDA	87	
Db	263	QGGDDGGEPSQLHPCHSVSPSAQPTWTHPCGOSQLFADLSREELTAVMKFLQRLGPGTVDA	322	
Qy	88	QARPSDNCVFSVELDLPKAAALAHLDKRSPPAREALAIFFGSGQPDVNSVELVGPDP	147	

D	b	323	QARSDCVCSVLEQLERPKAALAHNRGSPRARELALVFFGRQOPRVNSELVUGPLR	382
Q	y	148	HPSYMDVYVERHGCLPLRYHRRPVLFOEYLDIDOMIFNRELPOASGLLHHCFCYKRRGRN	207
D	b	383	HPSYMRBVYVERHCGPLRYHRRPVLFOEYLDIDOMIFNRELPOASGLLHHCFCYKRRGRN	442
Q	y	208	LVMTTTPARGLOSGDRATWGLYNYNSIGAARFLHHNGLELLVNNHKALDPAKMTIOKVFYO	267
D	b	443	LVMTTTPARGLOSGDRATWGLYNYNSIGAARFLHHNGLELLVNNHKALDPAKMTIOKVFYO	502
Q	y	268	GRYDLSLAOLEAGEADLVNWNVLIPDNGTSGWSLKSVPBPGRAPLQFVPOGPRFSVOC	337
D	b	503	GRYDLSLAOLEAGEADLVNWNVLIPDNGTSGWSLKSVPBPGRAPLQFVPOGPRFSVOC	562
Q	y	328	SRVASSLMTFSFGIGAASGRPIEDVPRROGERLYEISLOALALYGGNSPAAMTTRVYOG	387
D	b	563	SRVASSLMTFSFGIGAASGRPIEDVPRROGERLYEISLOALALYGGNSPAAMTTRVYOG	622
Q	y	388	GFGGKGYTTPLTRGVDCPRYLATVYDMHFLLESQAPKTIRAFCEFEONOGCLPLRRHNSDL	447
D	b	623	GFGGKGYTTPLTRGVDCPRYLATVYDMHFLLESQAPKTIRAFCEFEONOGCLPLRRHNSDL	682
Q	y	448	YSHFPGGLAEVLYVVRKSNSTLANTDYWMDYVFNHSGALIEIRPATGYISSAFLFGATGKY	507
D	b	683	YSHFPGGLAEVLYVVRKSNSTLANTDYWMDYVFNHSGALIEIRPATGYISSAFLFGATGKY	742
Q	y	508	GNQVSEHTLGTVHSHSAFKVDLDVAGLENNVMAEDVAFVPMVPMSPSPEHQLORLOYTRK	567
D	b	743	GNQVSEHTLGTVHSHSAFKVDLDVAGLENNVMAEDVAFVPMVPMSPSPEHQLORLOYTRK	802
Q	y	568	LLEMEBOAFLVGSATPRYLILASNSHNSKMGHPRGYRIQMLSFAGEBPLPONSSMARGFSW	627
D	b	803	LLEMEBOAFLVGSATPRYLILASNSHNSKMGHPRGYRIQMLSFAGEBPLPONSSMARGFSW	862
Q	y	628	ERYQLAYVQREKEEPSSSSVFNQNDPMAPIVYDSDFINNETIAGKDLVAAVTAGFLIHP	687
D	b	863	ERYQLAYVQREKEEPSSSSVFNQNDPMAPIVYDSDFINNETIAGKDLVAAVTAGFLIHP	922
Q	y	688	AEDIPNTVTVNGVGFELRPYNEFDEDPSEFYSADSIYFRGDODAGACEVNPPLACLPAAAA	747
D	b	923	AEDIPNTVTVNGVGFELRPYNEFDEDPSEFYSADSIYFRGDODAGACEVNPPLACLPAAAA	982
Q	y	748	CAPDLPAPSHGGSFHN 763	
D	b	983	CAPDLPAPSHGGSFHN 998	

RESULT 3
 US-09-919-497-51
 Sequence 51, Application US/099919497
 Patent No. US2002010662A1
 GENERAL INFORMATION:
 APPLICANT: Mutter, George L.
 TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
 FILE REFERENCE: B0801/7225
 CURRENT APPLICATION NUMBER: US/09/919,497
 CURRENT FILING DATE: 2001-07-31
 PRIOR APPLICATION NUMBER: US 60/221,735
 PRIOR FILING DATE: 2000-07-31
 NUMBER OF SEQ ID NOS: 100
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 51
 LENGTH: 729
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-919-497-51

Query Match	61.9%	Score 2336.5;	DB: 10;	Length 729;
Best Local Similarity	63.4%	Pred. No. 5.7e-215;		
Matches 483; Conservative	90;	Mismatches 154;	Indels 35;	Gaps 4;
OY	1	MNQKTIIVLLIAVTTITALVCVLVGRGDCGEPSQLPHCPVSVPDAQPMTHPGOSQLF	60	
	:	: : : : : : : : : : :	:	:

```

Db      1 MHKIVLAFALSLITIFLAVLLTSPGSGS-----SQPHCSVSHRAPNHPGSOQLF 56
Qy      61 ADLSREELTAVNRFLTORLQGLVDAQAQPSDNCVFSVELQPPKAAALHLDRSGSP 120
Db      57 ADLSREELTAVNRFLTORLQGLVDAQAQPSDNCVFSVELQPPKAAALHLDRSGSP 116
Qy      121 ABEALAIYFEGROPQPNVSELYVGPLPHPSYMRDVTVERHNGPLPHRRPVYFOELDD 180
Db      117 ABEALAIYFEGROPQPNVSELYVGPLPHPSYMRDVTVERHNGPLPHRRPVYFOELDD 176
Qy      181 OMIFNELDQAAGSLHHCCEFYKHGRNLVTMTAPRGLOSGDAPTEFGYUNISGAFPL 240
Db      177 RHLKVELPKARIFLST--FVINGSTLAAVHATPRGLSREKTTTIGLHINISGGL 234
Qy      241 HHVGLLELVNHRKAPARTIQKVFQGRYVDSLAQLEAFEGAGLVNVLIPDNGTGSN 300
Db      235 HPVGLLELLDHRALDPAHMTVOOVFLGHYVADLGQLEBEFKSGRLVVRVPLPPNGAS 294
Qy      301 SLKSPVPPGAPPLQFYPGQPRFSVSGSRVASSLMTFSGLGAFSGPRIFDVAFQGERLY 360
Db      295 SLRSRNSPGLPLQFSPGSOQSVQGNLVSSLSMSTFGHGVSGLRIFDVAFQGERLY 354
Qy      361 YEISLEALAIYGGNSPAAVTTRYVDGFGMGKTYTPPLARGVDCPYLATVVDHFLLESQ 420
Db      355 YEVSVECVSYIGADSPKTMILTRYLDSFGLGNSGLVARGVDCPYLATVVDHFLLESQ 414
Qy      421 APRTIRDAFCVEONOGDLRLRRHSDLYSHYEGGLAEVLYVNSMSTLNYDVMDTVF 480
Db      415 AVQLLEAGVAFVEEAGGLPLRRHNYLQNHFYGLASSALVYRSVSGYDYIMDFVLY 474
Qy      481 PGATIRRYRACYSISATLFGATG--KGNQVSEHTLGTVNHSHAFKYDLVAGLENN 538
Db      475 PNGAIEGRVHATGYINTALKGEGELLFGRNREVLGTVNHSHAFKYDLVAGLENN 534
Qy      539 VVAEDVYFPMAPVSPSEHOLRLQVTRKLEKEEOAALFVGSATPRYLYLASNHSKMG 598
Db      535 VVAEDVYFPMAPVSPSEHOLRLQVTRKLEKEEOAALFVGSATPRYLYLASNHSKMG 594
Qy      599 HPRGYTIQMLSFAGEPLPONSSNARGFSWERQIOLAVTQKKEEPSSSSVFNQDPAVY 658
Db      595 HPRGYTIQMLSFAGEPLPONSSNARGFSWERQIOLAVTQKKEEPSSSSVFNQDPAVY 654
Qy      659 DSDFTINNETIAGKDLVAMVTAGELIHPAEDIPNTVTGNGGCFELRPVNFEDDEPSIF 718
Db      628 TFAEDFTINNETIAGKDLVAMVTAGELIHPAEDIPNTVTGNGGCFELRPVNFEDDEPSIF 687
Qy      719 SADSIYFEGDODAGACEVNPPLACLPOAACAPDLPAFSGCF 760
Db      688 SPSGVYFEKGQDAGLCSINPVACLPDLACVPLPFPFSGYHGF 729

```

```

RESULT 4
US-10-005-530-58
; Sequence 58, Application US/10005530
; Publication No. US20030026795A1
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
; FILE REFERENCE: 11899.0022.DVUS01 (MOBT-022--2)
; CURRENT APPLICATION NUMBER: US/10/005,530
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/063,733
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/044,504
; PRIOR FILING DATE: 1997-04-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 617
; TYPE: PRT

```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (74)..(79)
; OTHER INFORMATION: Xaa - Unknown
US-10-005-530-58

```

```

Query Match          2.6%; Score 106; DB 9; Length 617;
Best Local Similarity 22.8%; Pred. No. 1.4;
Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;

```

```

Qy      63 LSREELTAVNRFLTORLG--PGLVDAQAQAR-----PSDNCVFSVELQPPKAA 109
Db      18 LIREKVAKEMLITERLGEVPGIPRPNNEGRFLGGYSHDNLPSDPIYSSTK----PGGGG 73
Qy      110 LAHLDRGSPPARAEALAIYFEGROPQPNVSELYVGV-----PLPHPSYMRDVTVERHNGP 163
Db      74 XXXXXGCGPPRK--VCIVAG-----VSGLIAMILDDKIPNLTYDIFESSRTGGR 125
Qy      164 LPHRRPVLYFOEYLDIDOMIFNELDQAAGSLHHCCEFYKHGRNLVTMTAPRGLOSCDR 223
Db      126 LYNHFTDAKHQYDIDGAMRY-----PDPSMKRTFNLFRKTGMLI-----KYLIDGEN 175
Qy      224 ATYFGLIYNISGAGFLVHVGLELVN-----HKALDPARTQKVFQGRYVDSLA 275
Db      176 TP--QLYNN--HFPAKGVSDPYNWSVANGCTVPDDVYDSVEKIQAF--GYREKLA 227
Qy      276 QLEAOFEGAGLVNVLIPDNGTGSWSLKSPPGAPPLQFYPGQPRFSVSGSRVASSLW 335
Db      228 E--DFDKGDEMLVDMDMT--REYLKRGPRGAEAKPYF-----FALQMMETON--- 273
Qy      336 TFSFGLGAFSGPRIFVRFQGERLYVEISLQ-----EALAIYGENS 376
Db      274 -----TGTNLFDOAF--SESVISDFDNDPTKPEWCIEGGS 309

```

```

RESULT 5
US-09-908-193-22
; Sequence 22, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHINKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-908-193-22

```


APPLICANT: IKEDA, MASATO
APPLICANT: OKAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5832
LENGTH: 1194
TYPE: PR1
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5832

Query Match 2.5%; Score 102; DB 9; Length 1194;
Best Local Similarity 18.9%; Pred. No. 8.5;
Matches 112; Conservative 71; Mismatches 207; Indels 204; Gaps 29;

QY 214 APRLOSGDRATWGLY---NISGAFELHVGLELLVNNKALDPAKRTIOKVFYQGRY 270
DB 80 APESRFKERRWGEPPKSDVSGAYLHOT---MLAENTT-----GLRNF---Y 126
QY 271 YDSIAOLEAOPFALVNVVLLIPDNGTSGWSLKSPPVPPAPPLQFYQGRFSVQGRV 330
DB 127 LSSNASTYEGOL-----GKW-----PRMD----- 144
QY 331 ASSMTSTSFGLGASGRPIEDVROGERLYEISLQELA---IYGNSSPAMTTRYDG 387
DB 145 ADIAEHAEGIIATTGCPGSDVQTRLRIGCFDEALEAAMQDIYGRDN---YELELDH 201
QY 388 GFG---MGKTTPLTGVCPCPLATYVDMHFLAESQAKTIRDAFCFEONOGPLRRH 443
DB 202 GLDIETRRSELEIGRLNLPPLVTN-DCHVLESQA--OAHMLCVOGKTL---H 254
QY 444 HSDLYSHYFGGLAETVLVVRSMSTL-----LNDYVMDTVFHPGSA 484
DB 255 DEDFK--FGGTGYVVSABQMRLAMDMPDGDONTLMIARQSYDEIMEE--HSHDR 310
QY 485 IEIPIFYATGYISSAFLEGATGKYNQVSEHTLGIVHTSHAFKYDLDVAGLENNVWADM 544
DB 311 MPIDVPEGYPTTWM-----HHEV---NAGLEDREFSGQOV 343
QY 545 VFVMAVMPSPHOLQRLQVTKLLEMEQOAPLVGSATPRYVLAS---NHSN---KM 597
DB 344 -----PEYIEREYEISVIYDMK-----GYSTFLVAIILIHANSIGLRV 384
QY 598 GHPGRIOMLSFAGEPLPQNSSMARQSWERYQLAVYQREEEPSSSSVFNQNDMPAP 656
DB 385 GPGGSAAGALVAYALTITNDPHEHGLPERF-----LNPERSPAPD 427
QY 657 -TVFSPDINNETHA-----GKDLVAVY-TAGFLIHAEIDIPNTVYVGVGFFLAPY 708
DB 428 IDIDFDDRRREMIRYAADRWGEKIAQVITFGYTKRQALKDSARVQMG----- 477
QY 709 NFEDDEPSFYADSIFYRGDDAGACEVNPPLAC---LPQAAACAPDLPAFSGHG 759
DB 478 -----QPGYQIADRY-----IKELPPAIMAKDIFLSGITDPDHRFENAG 517

RESULT 8
US-10-005-530-46
Sequence 46, Application US/10005530
Publication No. US20030026795A1
GENERAL INFORMATION:
APPLICANT: Isaac, Barbara G.
APPLICANT: Greenplate, John T.
APPLICANT: Purcell, John P.

APPLICANT: Romano, Charles P.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
FILE REFERENCE: 11899.0022.DVUS01 (MOBT:022--2)
CURRENT APPLICATION NUMBER: US/10/005,530
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: 09/063,733
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/044,504
PRIOR FILING DATE: 1997-04-21
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 46
LENGTH: 617
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Polypeptide
US-10-005-530-46

Query Match 2.5%; Score 101; DB 9; Length 617;
Best Local Similarity 23.3%; Pred. No. 3.8;
Matches 81; Conservative 39; Mismatches 138; Indels 90; Gaps 19;

QY 63 LSREELNAVMBFLNORIG--PGLVDAQAR-----PSDNCVSEVLQLPKRAA 109
DB 18 LIREKVAKEILNTERLGEVPGIPPNBGRFLGGYSHDNLPSDPLYSI-----KPAL 71
QY 110 LAHLDRGSP--PPAREALAIYFGRQOPQVNSVLVG-----PLPHSPYRDYVERHG 161
DB 72 LKEAPRAEELPPRRK--VCIVGAG-----VSGLYIAMIIDLDLKIPNLTYDIFESSRGT 123
QY 162 GPLPYHRRPVFOEYLDIDQIFNRELQASGLLHCCFYHRRGNLVMTARGLDGS 221
DB 124 GLTYHTHTDAKHYYDYGARY---PDIPSMKRTENLFRGTGMPIL-----KYILDG 173
QY 222 DRATWFGLYYNSGAFELHVGLELVN-----HKALDPAKRTIOKVFYQGRYDS 273
DB 174 ENTP--QLYNN-----HFFAKGVSDPYMVSANGGVDPDVYDSVGEKIQOAF--GYREK 225
QY 274 LAOLEAOPFALVNVVLLIPDNGTSGWSLKSPPVPPAPPLQFYQGRFSVQGRV 333
DB 226 LAE---DFDKGFDELMDDMTT--REYLKRGKPGKEAPKYDF-----FAIOWMETQN- 273
QY 334 LMTFSFGIAGSGRPIEDVROGERLYEISLQ-----EALAIYGNSS 376
DB 274 -----TGTLNLFQDAF--SESVIDSFDFDNFKPEMYCIEGTS 309

RESULT 9
US-10-213-990-27
Sequence 27, Application US/10213990
Publication No. US20030082595A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Bussey, Howard
APPLICANT: Bussey, Reg
APPLICANT: Roemer, Terry
TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL
FILE REFERENCE: 10182-019-999
CURRENT APPLICATION NUMBER: US/10/213,990
CURRENT FILING DATE: 2002-08-05
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 944
TYPE: PR1
ORGANISM: Aspergillus
US-10-213-990-27

Query Match 2.4%; Score 99; DB 9; Length 944;
Best Local Similarity 20.2%; Pred. No. 11;
Matches 122; Conservative 62; Mismatches 243; Indels 176; Gaps 29;

```

OY 174 OEYLDIDMIFRRELPOASGLL-----HHCFYKHRRRLVMTTAPRGLSGDRAT 225
DB 444 OYNLSISTSVGNVTIPQAGGSLSLNGRDSKPHITDIDVGFEPLITSSAEVFTWANGDNKK 503
OY 226 WFGLYYNISGAGFLLHVGLELVNHNKALDPAWMTQKVFYOGARYDSIAQLEAO---FE 282
DB 504 RVLVLY--GGAG-ELHEFALPKHLPRPTVEGSSV--KIAKGSAMVWQEWMAARVRLR 558
OY 283 AGLVNVVILPONGIGSSLSKSPVPPGAPPLQFY--PQGPFFSVQGSRVASSLW----- 335
DB 559 AGKLEIHLHMRNDAYOHVLELPA---KOPIANYSPPSKETVIKGGGLLSAMITDND 614
OY 336 -----TFSGGLAFSGPRIFD--VRFOGERLYVEISLOEALAIIGNSPAA----- 379
DB 615 LHLGADVAVTTPLEYISAPKRFEDGIVENGOSIKSTRKIGLATAVHQPPAISLPDLKR 674
OY 380 -----MTTRYVDGFGMGKRYTTPLTRGVDCPYLATYVDMHFLLESQAPKTIRDA 428
DB 675 LDMKYIDLPELSTFYNDEGM-----TPLTN-----TY-----TNNTRFPGP 712
OY 429 FCVEFQONGGLRRHNSDLVSHYFGGLAEVYLVRMSMTLLNYDYVMDTFHPSGAIEIR 488
DB 713 TOLYADYV-----YHGG-----SLIYRGHFT--ANGDESM--VFLNTSGG--VG 751
OY 489 FYATGYISSAFLEFGATGKYGNQVSEHTIGYVHTSAH-----FKYDLVAAGLENMTWADM 544
DB 752 FANSWMLQOTELGSWTSGGRNMTYPRNISLPHLSPGEPYFTVVIDHNGOEEAPGTDA 811
OY 545 VFVPAVWSPSEHOLQLOVTRKLEMEEOAFLVGSATPRLYLASNH---SNKW---G 598
DB 812 IKFP-----RGILD-----VALSGHELSDLKMKMG 837
OY 599 HPRGRITMLSFAGEPLPQNSSMARGFSWERYOLAVTORKEEPPSSSVFNONDPAWPTV 658
DB 838 NLGQGYODLLRG--PLNEGAMYAR---OGYHL-----PSPPTSS--WKSSNP----- 879
OY 659 DESDEINNETIAGKDLVAMVTAFLHPIHAEIDIPNTVTVNGCVG-----FELRP 707
DB 880 -----IHEGLTAGIGIYATSFSLDLPBGYDIPLSFPPNNSASARSGTIRCOLFVNG 933
OY 708 YNF 710
DB 934 YQF 936

RESULT 10
US-09-738-626-4914
: Sequence 4914, Application US/09738626
: Publication No. US20020197605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIAI, KEIKO
: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAOKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT FILING DATE: 2000-12-18
: PRIOR FILING DATE: 1999-12-16
: PRIOR FILING DATE: 1999-12-16
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
```

```

: SEQ ID NO 4914
: LENGTH: 581
: TYPE: PRT
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-4914
Query Match 2.3% Score 95.5; DB 9; Length 581;
Best Local Similarity 22.8%; Pred. No. 11;
Matches 115; Conservative 50; Mismatches 188; Indels 151; Gaps 29;

OY 260 TIQKRYOGRYDSIAQLEA--OFEAGLVNVV-----LIPDNGT--GGSMSLSKSPV 307
DB 164 TITKVIHDDGTEEGLAQLALOFETKQVDYEDLYRETRASLTPEDQITLAPRWGVATPT 223
OY 308 PG---PAPLPQFYPGCPNPSVQGSRVASSLWTFSGGLAFSGPRIF-----DYRF 354
DB 224 SDDKLAAPPLQSKRDGEMSPIV---IASP---ESGISAVATARAFAGAIREMDYDPDRF 276
OY 355 OGERLYVEISLOEALAIIGNSPAAWMTTRYVDGFGMGKRYTTPLTRGVDCPYLATYVDMH 414
DB 277 NOEQKMYAGLADRLPLALDQ-----FGTG----- 302
OY 415 FLLESQAPKTIRDAFCVEONONG-----LPLRRHSDLYSH---YFGGLAE---TVLYVR 463
DB 303 ---EQLAEKIRKEEVTTELPGGGLVFPGRMVA--LIGHPSPALGALGEPAPAEAVAR 358
OY 464 SKSTLLNTDYVDVTFHPSGAIEIRFYATGYISSAFLEFGATGKYGNQVS--EHTLGTVH-- 520
DB 359 ATDMAOQOEFEQDQVIPA-----FEVIAVVASBFA--GDDQNTSENEPVEDLVGYDAI 411
OY 521 THSAFXYVDLV--AGLEKMWNAEDVFPMAVPSPEHOLQLOVTRKL-----LEMEEO 574
DB 412 TEAGYAV--LDLQGRANFLQAKKY-----EELKRPNGVGLALDPEMKIGPQEO 460
OY 575 AAFIVSATPRLYLASNHKMGHPRGYRIOMLSFAEPLPQNSSMARGFSWERYOLAV 634
DB 461 PMTRVGSADAELI-----NEVSQM-----LADL--AENDLPO-----KATVLAHQFLEM 503
OY 635 TORKEEPPSSSVFNONDPAWPTVPSDFINNETIAGKDLVAMVTAAGFLHPIHAEIDINT 694
DB 504 LTRNDQ-----INTDLELAVLVHADGHN--AGDKYATW-----NM 538
OY 695 VTVNGVGFELRPYNEFDED--PSF 717
DB 539 LREGLSPNITFMAKNFYDEDPTE 562

RESULT 11
US-10-005-530-57
: Sequence 57, Application US/10005530
: Publication No. US20030026795A1
: GENERAL INFORMATION:
: APPLICANT: Isaac, Barbara G.
: APPLICANT: Greenplate, John T.
: APPLICANT: Purcell, John P.
: APPLICANT: Romano, Charles P.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
: FILE REFERENCE: 11899.0022.DVUS01 (MOBT:022--2)
: CURRENT FILING DATE: 2001-10-26
: PRIOR FILING DATE: 2001-10-26
: PRIOR APPLICATION NUMBER: 09/063,733
: PRIOR FILING DATE: 1998-04-21
: PRIOR APPLICATION NUMBER: 60/044,504
: PRIOR FILING DATE: 1997-04-21
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 57
: LENGTH: 617
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic Polypeptide
: NAME/KEY: MISC_FEATURE
```

LOCATION: (74)..(79)
OTHER INFORMATION: Xaa = Unknown
US-10-005-530-57

Query Match 2.3%; Score 95; DB 9; Length 617;
Best Local Similarity 22.5%; Pred. No. 13;
Matches 78; Conservative 40; Mismatches 142; Indels 86; Gaps 18;

QY 63 LREBELTAVMRELTORLG--PGLVDAQAAR-----PSDNCVFSVEIQLPPKAA 109
DB 18 LIREKAKELNLTIRLGEVPGIPRPNRGRFLGGVSHNDLPDLSSTI-----GGSGG 73
QY 110 LAHLBGSPPRAREALATYFFGHOPNPNSLYVG-----PLPHSYARDYVERHGGP 163
DB 74 XXXXXGGGPPRR--VCYIAG--VSGLYAMILDCLKIPNLTIDFESSSRGTGR 125
QY 164 LPHRRPVLFOEYLDIDOMIFNRELPOASGLHCCFYKRGHNLVTMTAPRGLQSDR 223
DB 126 LTHHTDAKHXYDIGMRY---PDISMKTENLFRKTGMPLI-----KYYLDGEN 175
QY 224 AMWGLYVVISGAGFLHVGLELVN-----HKALDPAKWTQKVYQGRYDSLA 275
DB 176 TP--QLYNN---HFFAKGSDPYMVSANGTVPDDVDSVEKIQOAF--GYREKLA 227
QY 276 QLEAOFEAGLVNVLIPDNGTGSMSLSPVPPGAPPLDFOYPOGRFVSQSRVASSLM 335
DB 228 E--DDEKGFDELMLVDMMT--REYLKRGKGPKEAPKYDE-----FALQMETON-- 273
QY 336 TSEFGAFAFGPRIFDVRFOGERLYEISLQ-----EALAIYGSN 376
DB 274 -----TGTWLFDOAF--SESVIDSPEDNFTKPEMTCIEGSTS 309

RESULT 12

US-09-912-020-364
Sequence 364, Application US/09912020
Patent No. US20020045592A1
GENERAL INFORMATION:
APPLICANT: zyskind, Judith
APPLICANT: Ohlsen, Karl L.
APPLICANT: Trawick, John
APPLICANT: Forsyth, R. Allyn
APPLICANT: Froelich, Jamie M.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
FILE OF INVENTION: ESCHERICHIA COLI
FILE REFERENCE: ELITRA 001DVI
CURRENT APPLICATION NUMBER: US/09/912, 020
CURRENT FILING DATE: 2001-07-23
PRIOR APPLICATION NUMBER: 09/492, 709
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: 60/117, 405
PRIOR FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 485
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 364
LENGTH: 878
TYPE: PRF
ORGANISM: E. Coli
US-09-912-020-364

Query Match 2.3%; Score 93.5; DB 10; Length 878;
Best Local Similarity 20.2%; Pred. No. 30;
Matches 108; Conservative 61; Mismatches 190; Indels 175; Gaps 27;

QY 176 YLDIDOMIFNRELPOA-----SGLLHCCFYKRGHNLVTMT-- 212
DB 148 HLDVGGQRLNLTIPQAFMSNRARGYIPPELMDPGINAGLLN---YNSGNSVONRIGN 203
QY 213 --TAPRLOGS-----DRATWFGLYN---ISGAGFLHVGLELVNKKALDP- 256

DB 204 SHKATVLOSGLINQAMRLNDNTM---SYNSDRSSGSKNKKQHINTML---ERDITPL 257
QY 257 -ARMTIQKVFYQGRYDSL---AOLEAOFEAGLVNVLIPDNGT----- 297
DB 258 RSRULTGGDYTGQDIPDGINFREGAQLASDNN-----MLPDSORGFAPVINGIARGTAG 310
QY 298 -----GMSLSKPVPPGAPPLQATFYPOGRFVSQ-----GS-----R 329
DB 311 VTIKONGYDIYNSTVPPGFTINDIYAAGNSGDLQVTKHEADSGTIFTPVSSVPLOR 370
QY 330 VASSLMTSEFGAFAFGPRIFDVRFOGERLYEISLOEALATIGNSPAAMTTRYVDGF 389
DB 371 EGHTRISTAGEYRSGNAOQEKTRFFOSTLH--GLPAGTITYGTO--LADRYAFNF 425
QY 390 GMKRYTTPPLRGVDCPYLATYVDMHFLLESQAKTIRDAFCVEONOGILPLRRHSDLYS 449
DB 426 GICKNMGALG-----ALSDM-----TOANSTLPDD-----SOHDGQSVR---FLYN 464
QY 450 HYFGGLAEVLYVRSMTLLNIDYVMDTVPH-----PSGAIEIRFATGYISSAFL 500
DB 465 KSLNESGTNIQLVGRYSTISGYFNADTTYSRMNGINIEFQDVIQKPKFTYINDA-- 522
QY 501 FGATGKYGNQVSEHTLG--TVHTSHAH-----FKYDLVAGLENNVMAEDMV 545
DB 523 YNRKGLQLTVTQO--LGRSTLYLSGSHQTYMGTSNVDQFOABLNTA-----FED-- 572
QY 546 FVPMAPVMSPEHOLQRLQVTRKILLEMEQAPLVGSATPRYLYLASHNSKKMGH 599
DB 573 -----INWLTYSYL-----TKNAMQKGRDOMLALNVPISPHWLRSDSKQSMRH 616

RESULT 13

US-09-734-569-172
Sequence 172, Application US/09734569
Patent No. US20020064816A1
GENERAL INFORMATION:
APPLICANT: Lersch, Jens
APPLICANT: Renz, Andreas
APPLICANT: Ehrhardt, Thomas
APPLICANT: Reindl, Andreas
APPLICANT: Cirpus, Petra
APPLICANT: Bischoff, Friedrich
APPLICANT: Frank, Markus
APPLICANT: Freund, Annette
APPLICANT: Duwenig, Elke
APPLICANT: Schmidt, Ralf-Michael
APPLICANT: Reski, Ralf
TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins invc
FILE OF INVENTION: in the synthesis of carbohydrates
FILE REFERENCE: BASF-NAE-1332-99-US
CURRENT APPLICATION NUMBER: US/09/734, 569
CURRENT FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: US 60/171, 101
PRIOR FILING DATE: 1999-12-16
NUMBER OF SEQ ID NOS: 181
SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1
SEQ ID NO 172
LENGTH: 480
TYPE: PRF
ORGANISM: Physcomitrella patens
US-09-734-569-172

Query Match 2.3%; Score 93; DB 10; Length 480;
Best Local Similarity 22.9%; Pred. No. 13;
Matches 91; Conservative 55; Mismatches 124; Indels 128; Gaps 22;

QY 1 MNQTIYLVLLAVI-----TIFALVCY--LLVGRGDDGEGEPQLPHCSYS 45
DB 168 LNDLFCVIGVILLADPLGNSNPMWRTYFALATVAVLGLD-----MAYCPE-S 218
QY 46 PSAQPTWHPGOSQLFADLSRELTAVMRFLTORLGPGLVDAQAARPSDNCVFSVEIQLDP 105
DB 219 PR---WLXKXK-----TAEATVAVRL---WGAKEVSSMADLAASSVETVK----- 260

QY 106 KAALAHLDGRSPPPAREALAIFFGROPQPNVSELYVGLPHPSYMDVVERHGGPLP 165
Db 261 -----GDTQDASWGLFGRK-----YRKVVTV---GMAL- 286
QY 166 YHRRPVLEOYLIDQMI-FNRELPOASGLHHCCFYKHGRNLTMTARGL--QSGD 222
Db 287 -----FLEOGFAGINAVYFSTGYFRSAGITNDVAASALVAGAVAGTTVASGMMDKGR 341
QY 223 RAYFGLYXNSGAGFLHHVGLLELVNHNKALDPARMTIQVFYQGRYDSLAOLEAOFE 282
Db 342 KSLMG-----SFAGMSLIMLVLSLWSPDLA-----YSGTIAVL----- 378
QY 283 AGLVNVLIPDNGSGMSLSKSPVPGPAPPLQFYPO--GPRFSVQGSRAVSSU---WTF 337
Db 379 -GTYSYL-----SFSIGA---GPVREL-LLEPLFGAR--IRAKAVALSLGVMWC 422
QY 338 SFGIAGFSGPRIFDVRFQGERLYVEISLQALAIYGCN 375
Db 423 NFMIGLEFLNVLVQKFGVSTVYLFPSAVCAAIAYVGN 460

RESULT 14
US-09-925-301-844
; Sequence 844, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925.301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 844
; LENGTH: 601
; TYPE: PR
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (36)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (64)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (103)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (152)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (358)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (383)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-844

Query Match 2.38; Score 93; DB 10; Length 601;
Best Local Similarity 22.3%; Pred. No. 19;
Matches 78; Conservative 32; Mismatches 112; Indels 128; Gaps 19;
QY 31 DGGSPQLPNC-----PSVSPSA-QPWTHP-----GOSLFLDLSREE 67
Db 218 DTGTEALTPHIMRNLRHATSRKSTYRPSSEMERWEPILSPFEDVAGTEMSQSDSGVLDGDS 277

QY 68 LTAVMRFLTORLGP--GLYDAQAARP-----SDNCVSEYELQPPKAAALAHLDGRSPPA 121
Db 278 QVS--SGPCSQRSSPDGGLKGALEGPPKRPQSSPLNAPCEGPP-----GSEPPR 326
QY 122 REALAIVFGROPQPNVSELYVGLPHPSYMDVTVR--HGGPL-PYHR--RPVLF--- 173
Db 327 RPPAPHDGDKELPREOPPLPPPI---GTRSOXTRDGTGPIRSHRGRPPVQGTGX 383
QY 174 QEYLDI-----DOMIFNRE-----LPQASGLHHCCFYKHGRNLTMT 211
Db 384 DKSDRLRVVGDLSLKAERELTASVTEAIPVSRDMELLPSAASA-----EPOSKNIDSG 437
QY 212 TTAARGDSGRATWFGLYNISAGFELHHVGLLELVNHNKALDPARMTIQVFYQGRY 271
Db 438 HCVEPSSSGR-----LYPEVFF 456
QY 272 DSLAQLEAGFAGLVNVLIPDNG--TGGSMSL-----KSP--VPPGPAP 313
Db 457 GSAPSSSQISGAMDQLHFNHSGFRGTRPSLHPRYSQPLYLPPGPAP 506

RESULT 15
US-08-954-701A-19
; Sequence 19, Application US/08954701A
; Publication No. US20030032085A1
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley, Hoag & Eliot
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,701A
; FILING DATE: 20-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36709
; REFERENCE/DOCKET NUMBER: SUV-003.08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-954-701A-19

Query Match 2.38; Score 93; DB 1; Length 1447;
Best Local Similarity 25.5%; Pred. No. 71;
Matches 49; Conservative 28; Mismatches 67; Indels 48; Gaps 10;
QY 6 ILVLLIIVITIFLVGVLL--VGRGSDGGEPSQLPNCPSVSPSAOP-----WTH 53
Db 1156 LAITITIGVLLGLVLLVLSFRCPIPEVSPANGLNLRPTSPSPPPSVVRFAMPQETH 1215
QY 54 PGOSQLFADLSREE---LTAVMRFLTORLGP-----LYDAQAARPSDNCVSEYELQ 103

```
Db 1216 SGSDSDSEYSSQTVSGLSSELRLHYEAQGAGCPAHQVIYEA-----TENPVFAHSTVY 1270
OY 104 PPKA-----ALAHLDRGSPPPAREALATVFGROPQPN-VSELYVGPLPHPSYMD 154
Db 1271 HPESRHHPPSNPROQPHLDSGSLPPGRQ-----GQPRRDPPREGLWPPLYRP--RRD 1321
OY 155 ---VVERHGRP 163
Db 1322 AFEISTEGHGP 1333
```

Search completed: May 20, 2003, 12:02:20
Job time : 127.567 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:09:54 ; Search time 29.4394 Seconds
(without alignments)
2491.578 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095

Sequence: 1 MNOKTILVLLAVITTFAL.....QAAACAPDLPAFSHGFSHN 763

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR.73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4095	100.0	763	2	JC5234
2	3334	81.4	762	2	AS4411
3	1400.5	34.2	746	2	S34656
4	1394	34.0	751	2	AS4053
5	413	10.1	755	2	B41836
6	397.5	9.7	757	2	E64889
7	377	9.2	684	2	AS6102
8	368.5	9.0	660	2	G90330
9	363	8.9	648	2	A48646
10	357.5	8.7	756	2	T48139
11	353.5	8.6	650	2	G74412
12	340	8.3	638	2	JC2139
13	332	8.1	674	2	JC7251
14	327	8.0	687	2	T47403
15	320	7.8	674	2	C44239
16	302.5	7.4	759	2	E64854
17	300	7.3	660	2	AM2234
18	295	7.2	692	2	S04963
19	247	6.0	794	2	T39171
20	246.5	6.0	300	2	T48138
21	229	5.6	587	2	S21139
22	214.5	5.2	670	2	S71320
23	126	3.1	862	2	T46289
24	125.5	3.1	460	2	T48137
25	125	3.1	757	2	C84120
26	111	2.7	1322	2	D82685
27	107	2.6	435	2	A42672
28	106.5	2.6	896	2	F96523
29	106	2.6	3938	2	T42761

30	103	2.5	2224	1	KRU05	coagulation factor
31	102.5	2.5	591	2	C48364	cytochrome-c oxida
32	101.5	2.5	546	2	B75573	conserved hypothet
33	101.5	2.5	848	2	C65083	hypothetical prote
34	101	2.5	626	2	T35669	hypothetical ATP/G
35	100.5	2.5	597	2	C67381	sensor histidine k
36	100.5	2.5	654	2	S76870	hypothetical prote
37	100.5	2.5	1257	1	A41060	neural cell adhesi
38	100.5	2.5	1544	2	E59431	phosphoinositide-b
39	100	2.4	919	2	S42842	T16G12.2 protein -
40	99.5	2.4	579	2	E83144	hypothetical prote
41	99.5	2.4	2352	2	C83229	probable non-ribo
42	99	2.4	602	2	T35782	probable secreted
43	99	2.4	915	2	T03589	probable aspartate
44	99	2.4	2055	2	T00093	hypothetical prote
45	98.5	2.4	881	2	S25445	nitrate reductase

ALIGNMENTS

RESULT 1

JC5234
amine oxidase (copper-containing) (EC 1.4.3.6) AOC2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 18-Feb-2000
C:Accession: JC5234
R:Zhang, X.; McIntire, W. S.
Gene 179, 279-286, 1996
A:Title: Cloning and sequencing of a copper-containing, topa quinone-containing mon
A:Reference number: JC5234; M01D:97128319; PMID:9972912
A:Contents: Placenta
A:Accession: JC5234
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-763 <ZNA>
A:Cross-references: GB:U039447; NID:g1399031; PID:NAC50919.1; PID:g1399032
C:Comment: This enzyme catalyzes the oxidation of primary amines to the correspondi
Cu(II) and 1 molecule of covalently-bound topa quinone. It also can scavenge circu
C:Genetics:
A:Gene: GDB:AOC2; DAO2
A:Cross-references: GDB:4562632
A:Map position: 17q21-17q21
C:Superfamily: amiloride-binding protein
C:Keywords: copper; metalloprotein; oxidoreductase; quinoprotein; topaquinone
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-763/Product: amine oxidase (copper-containing) #status predicted <MAT>
F:444-520-522/Binding site: copper (His) #status predicted
F:471/Modified site: topaquinone (Tyr) #status predicted

Query Match 100.0%; Score 4095; DB 2; Length 763;

Best Local Similarity 100.0%; Pred. No. 1.5e-316;
Matches 763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNOKTILVLLAVITTFALVGVLLVGRGGGSEPSQLPHCPSEVSASQAPWTHPGSQSLF	60
DB	1	MNOKTILVLLAVITTFALVGVLLVGRGGGSEPSQLPHCPSPVSPAPWTHPGSQSLF	60
QY	61	ADLSRELTAVMRFLLQRLGPGVLDAAQARPSNCVFSVELQLPKAAALAHNDRGSPP	120
DB	61	ADLSRELTAVMRFLLQRLGPGVLDAAQARPSNCVFSVELQLPKAAALAHNDRGSPP	120
QY	121	AREALAIFFGRQPPNSELVYGPLPHPSYMDYTVERRHGGLPHRRPVLFQEXLDD	180
DB	121	AREALAIFFGRQPPNSELVYGPLPHPSYMDYTVERRHGGLPHRRPVLFQEXLDD	180
QY	181	OMTFNRELPOASGLLHCCFYKRGGRNLVTMTAPRGLOGDRATVFGLYNISGAGFL	240
DB	181	OMTFNRELPOASGLLHCCFYKRGGRNLVTMTAPRGLOGDRATVFGLYNISGAGFL	240
QY	241	HHVGEELLVNHKALDPAKRTIQKVFQGRYYDSLAOLEAFGLVNVVLIIPNGTGGSM	300
DB	241	HHVGEELLVNHKALDPAKRTIQKVFQGRYYDSLAOLEAFGLVNVVLIIPNGTGGSM	300

Qy	301	SLKSVPPGPAPPLDLOFYQOGRFSYQGSRVASLMTFSEFGATSGPPIFDVVRQGBEYV	360
Dd	301	SLKSVPPGPAPPLDLOFYQOGRFSYQGSRVASLMTFSEFGATSGPPIFDVVRQGBEYV	360
Qy	361	YEISLQELALAIYGGNSPAAMTTRVYDGGFGMGKYYTTLPRTRGVDPYLATVYDMHFLEEQ	420
Dd	361	YEISLQELALAIYGGNSPAAMTTRVYDGGFGMGKYYTTLPRTRGVDPYLATVYDMHFLEEQ	420
Qy	421	APKTRDQAFYCFEONOGILPLRRHSDLYSHFGGLAEYLVYVRKSTLNTDYWDYVFH	480
Dd	421	APKTRDQAFYCFEONOGILPLRRHSDLYSHFGGLAEYLVYVRKSTLNTDYWDYVFH	480
Qy	481	PSGAIEIRFYATGYISAFLEFGATGKYNGQVSEHTLGVTHSHAFFKVDLVAGLENNWV	540
Dd	481	PSGAIEIRFYATGYISAFLEFGATGKYNGQVSEHTLGVTHSHAFFKVDLVAGLENNWV	540
Qy	541	AEDWYFVPMAYPWSPEHOLORLQVTRKLLMEDEQAFLVGSATPRYLYLASNHSKMGHP	600
Dd	541	AEDWYFVPMAYPWSPEHOLORLQVTRKLLMEDEQAFLVGSATPRYLYLASNHSKMGHP	600
Qy	601	RGYRLQMLSPFGEPLPPONSSMARGFSMERYOLATYORKKEEPPSSSVYNODMPAPYDVF	660
Dd	601	RGYRLQMLSPFGEPLPPONSSMARGFSMERYOLATYORKKEEPPSSSVYNODMPAPYDVF	660
Qy	661	SDFINNETIAGKDLVAMVYTAGFLHIPHAEDIPTNTVYNGVGYFLRPNYFNEDEPSEFYSA	720
Dd	661	SDFINNETIAGKDLVAMVYTAGFLHIPHAEDIPTNTVYNGVGYFLRPNYFNEDEPSEFYSA	720
Qy	721	DSIYFRGPDQAGACVNPPLACLPPQAAACAPDLPASFHSGGFSHN	763
Dd	721	DSIYFRGPDQAGACVNPPLACLPPQAAACAPDLPASFHSGGFSHN	763

RESULT 2

amine oxidase copper-containing (EC 1.4.3.6), serum, precursor - bovine
C.Species: Bos primigenius tauros (cattle)
C.Date: 09-Sep-1994 #sequence, revision 06-Feb-1995 #text, change 18-Feb-2000
A.Accession: A54411; B38081; A48242; S65408
R.Mu. D.; Medzhidskiy, K.F.; Adams, G.W.; Mayer, P.; Hines, W.M.; Burlingame, A.L.; Smith, J.P.
J. Biol. Chem. 269, 9926-9932, 1994
A.Title: Primary structures for a mammalian cellular and serum copper amine oxidase.
A.Reference number: A54411; MUID:94193686; PMID:8144587
A.Accession: A54411
A.Molecule type: mRNA
A.Residues: 1-762 <MAN>
A.Cross-references: GB:S659583; NID:9546215; PID:RAB30397.1; PID:9546216
R.Mu. D.; Jones, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Kilmann, J.P.
J. Biol. Chem. 267, 7979-7982, 1992
A.Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine oxidase.
A.Reference number: A38081; MUID:92235001; PMID:1569055
A.Accession: B38081
A.Molecule type: protein
A.Residues: 463-465 'D', 467-473 'X', 475-485 <MUD>
R.Janes, S.M.; Mu. D.; Memmer, D.; Smith, A.J.; Kaur, S.; Maltby, D.; Burlingame, A.L.;
Science 248, 981-987, 1990
A.Title: A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of
A.Reference number: A48242; MUID:90260648; PMID:2111581
A.Accession: A48242
A.Status: preliminary
A.Molecule type: protein
A.Residues: 468-472 <MAN>
R.de Blase, D.; Apostolnik, E.; de Matteis, G.; Mondévi, B.; Morpurgo, L.
Eur. J. Biochem. 237, 93-99, 1996
A.Title: Half-of-the-sites reactivity of bovine serum amine oxidase. Reactivity and chemical
A.Reference number: S65408; MUID:96205913; PMID:8620899
A.Accession: S65408
A.Molecule type: protein
A.Residues: 463-469, 'X', 471-487 <DEB>
C.Superfamily: amiloride-binding protein
C.Keywords: copper; glycoprotein; oxidoreductase; quinoprotein; topaquinone
F11-16/Domain: signal sequence #status predicted <StG>

F.117-762/Product: amine oxidase (copper-containing) #status predicted <MAT>
F.136, 231, 665/Binding site: carbohydrate (Asn) (covalent) #status predicted
F.470/Modified site: topaquinone (Tyr) #status experimental

Query Match	81.4%	Score 3334	DB 2	Length 762
Best Local Similarity	80.6%	Pred. No. 3.6e-256		
Matches 610; Conservative	61;	Mismatches 76;	Indels 10;	Gaps 2

OY	14	VITTAALVCVLLNCR - -GGDGE-----PSOLPHCPSPSPSAQPMTHNGOSLFDL	63
Db	3	I F I F S L M T L L V M G R E E G V S E G V K Q C H S L P R C P S R S P S O P M T H D S O L F A D L	62
OY	64	SREELTAVMRELTOCLGGLVDAAOARPPSDNCVFSEVLQLPKKAALAHLDGRSPPARE	123
Db	63	SREELTTVMSELTOQLGGLVDLDAOARPSDNCVSVLEQLPCKAALAHLDGRSPPARE	122
OY	124	ALAIVFGRQOPNVSELVGPRLPHPSYMRDVYERHGGPLPYHRPVLFOEYLDIDOMI	183
Db	123	ALAIYVFEGGQPQPNTELVGPRLPQPSYMRDVYERHGGPLPYRRPVLREYLDIDOMI	182
OY	184	FNRLEPOASGLLHHCCFFKHGHGRNLVTMTTPARGLOSDRATWGLNYINISAGCFULHV	243
Db	183	FNRLEPOAGAYLHHCCSKYGOGOKLTLTNNSAPRGVQSDRSRTWFETIYNITIKGGYHLPV	242
OY	244	GLELLVNHKALDPAMWTLOKFVYOGRYYDSLQAOLEAGEAGLVNVLPRDNCTGGSMSLK	303
Db	243	GLELLVHKALDPAMWTLOKFVFGORYTENLAQLEBOCEAGOVNMYVLPDGTGTFMSLK	302
OY	304	SPVPBPAPPLQIFYPOGGRFSVOGSRVASLMTSESGLGATSGPRIFDFRGOGERLYEI	363
Db	303	SOVPBPAPPLQIFHPQGGRFSVOGNRVASSLMTFSFLGATSGPRAVDFVRGOGERLAYEI	362
OY	364	SLQELMALYYGGSNPAAMTTRTYVDGFGMGKTPTPLTRGVDCPYLATYVDHMFLESQAPK	423
Db	363	SLQELGAAYYGSTPAAMLTTRYMSGFGMGYPATPLIRGVDCPYLATYVDHMDHVVESOTPK	422
OY	424	TIRDAFCVFEONOGPLRRHNSDLXSHYFEGGLAEVLVLVRSMSLTLANTDYWDVTFHPSG	483
Db	423	TLHDFAFCVFEONKGPLRRHNSDLXSHTFPGVAQVLVLFRSYSTMLANTDYWDVTFYFNG	482
OY	484	AIEIRFYATGYISSAFLEGATGAKYGNQVSEHTLGTVHTSHAHFVKYLDVLAGLENMVMAED	543
Db	483	AIEVLTHTATGYISSAFLEGAAARYGNQVGEHTLGTVHTSHAHYKYLDVGLGENVMMAED	542
OY	544	MVFVMAVMPMSPEHLOLOQTARKLLEMEDQAAFVLSGATPRYLYLASHSKMKCHPGXY	603
Db	543	MAVFVTAIPWSPENHIORLOYTRKOLETEEDQAAFLGGASPRYLYLASKOSKKMHGHPGX	602
OY	604	RLOMSFPAZEPLOPSSNAARGSMERYOLATYORKEESEPSSSYVNONDMPAPTDFSDF	663
Db	603	RLOYSFPAZGPPPOPSNEBRASFMSGRYOLATIQRKETERPSSSYVNONDPMPTPIYDESDF	662
OY	664	INNETHIACKDLVAWYTJAGFLHIPHAEDIPTNTYVNGVGFFLARPYNFEDEDPSFYASADI	723
Db	663	INNETHIACKDLVAWYTJAGFLHIPHAEDIPTNTYVNGVGFFLARPYNFDQEPSMDSADSI	722
OY	724	YFRGODAGACEVNPPLACLPOAAACAPDLPFASHGF 760	
Db	723	YFRGODAGSCIEINPLACLPOAATCAPDLPVFSHGXY 759	

RESULT 3

amine oxidase (copper-containing) (EC 1.4.3.6), kidney, precursor - rat
N:Alternate names: amiloride-binding protein, long form
N:Contains: amiloride-binding protein, short form
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence-revision 13-Jan-1995 #text-change 18-Feb-2000
C:Accession: S36847; S36848; I51904; S34656; S34657
C:Linkage: E. Renard, S. Volley, N. Waldmann, R. Chassande, O. Ladunski, M.
Ehr. J. Biochem. 216, 679-687, 1993
A:Title: Molecular cloning and functional expression of different molecular forms of
A:Reference number: S36847; MUID:93387321; PMID:8375402

A:Accession: S36847
 A:Molecule type: mRNA
 A:Residues: 1-746 <LIN2>
 A:Cross-References: EMBL:X73911; NID:g395064; PIDN:CAA52116.1; PID:g395065
 A:Accession: S36848
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 529-746 <LIN2>
 A:Cross-References: EMBL:X73912; NID:g395066; PIDN:CAA52117.1; PID:g395067
 A:Title: Isolation of a rat amiloride-binding protein cDNA clone: tissue distribution an
 A:Reference number: 151904; MUID:94295632; PMID:8023885
 A:Accession: I51904
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 530-586, 'r', 588-632 <VER>
 A:Cross-References: GB:S70383; NID:g547214; PIDN:AA831157.1; PID:g547215
 A:Note: In Genbank entry S70383, release 116.0, the source is designated as Rattus sp.
 C:Superfamily: amiloride-binding protein
 C:Keywords: alternative initiators; oxidoreductase
 F:1-746/Product: amine oxidase (copper-containing), kidney, precursor #status predicted
 F:529-746/Product: amiloride-binding protein short form #status predicted <SHRT>

Query Match 34.2%; Score 1400.5; DB 2; Length 746;
 Best Local Similarity 39.1%; Pred. No. 9.1e-103;
 Matches 300; Conservative 131; Mismatches 280; Indels 57; Gaps 14;

18 FALVCLVLRGCGDGEPEQLPHCPSPVSAQPFTHPGOSLPADLSRELFVAVMFLNQ 77
 1 FALVCLVLRGCGDGEPEQLPHCPSPVSAQPFTHPGOSLPADLSRELFVAVMFLNQ 77
 5 FGMAATVILVQLQYDTASAVRTPE-----DKARVADLSPODKAVHSFLNM 50
 78 RLPGGLVDAQAARSDNCVFSVELQLPRAAALAHLDKSGPPAREALAVFEGRQPN 137
 1 RLPGGLVDAQAARSDNCVFSVELQLPRAAALAHLDKSGPPAREALAVFEGRQPN 137
 51 REELGQPSKEPTLAKNSFLLEMLPKKKHVKFLDEGKGFNRRAAVIFPGADDPN 110
 1 REELGQPSKEPTLAKNSFLLEMLPKKKHVKFLDEGKGFNRRAAVIFPGADDPN 110
 138 VSELVYGPLPHPSYMRDVERHGSGPLPYHRRVPLFOEYLDIDQMIFNRELPGASGLHH 197
 1 VSELVYGPLPHPSYMRDVERHGSGPLPYHRRVPLFOEYLDIDQMIFNRELPGASGLHH 197
 111 VTFEAVGPLPRPYTRALTS-PRGHHLSWSSRPISTAEY---DLVTH-TLKRAITPLHQ 164
 1 VTFEAVGPLPRPYTRALTS-PRGHHLSWSSRPISTAEY---DLVTH-TLKRAITPLHQ 164
 198 -----CCEYKHKGRNLVMTTAPRGLOSGDRATWFGLYYNSGAFELHNGLELVN 250
 1 FLDLTGFSFGLGDDRCILFTDVAAPRGVASGQGRSMFVQRYVE--GYFLHPRGLEILD 222
 165 FLDLTGFSFGLGDDRCILFTDVAAPRGVASGQGRSMFVQRYVE--GYFLHPRGLEILD 222
 251 HKAIDPARMTQKVFYQGRYDSLAOLEAOFEGALVNVVLIPD---NGTGSNSLSKSPV- 306
 1 HKAIDPARMTQKVFYQGRYDSLAOLEAOFEGALVNVVLIPD---NGTGSNSLSKSPV- 306
 223 HGSTDQVDRKVEQLMWNGKFLYNNPEELARKYAVGEVDTVLEDPPLPNGTE----KPLPL 277
 1 HGSTDQVDRKVEQLMWNGKFLYNNPEELARKYAVGEVDTVLEDPPLPNGTE----KPLPL 277
 307 -----PPGAPRPLQFYPGGRFVSQGSRVASLSLTFEFGAGFSGRPIITFYVRQ 355
 1 PPGAPRPLQFYPGGRFVSQGSRVASLSLTFEFGAGFSGRPIITFYVRQ 355
 278 SSTKPRGEHTFPVNAVGPVAVGSGPRYKLEGMTVLXGGSFYRLKSSSGLOIFNVLEFG 337
 1 SSTKPRGEHTFPVNAVGPVAVGSGPRYKLEGMTVLXGGSFYRLKSSSGLOIFNVLEFG 337
 356 GERLYVEISLOEALATYGNSPAAMTTRYDGGFGMGKTYPTPLTRGVDCYALATYVDMHF 415
 1 GERLYVEISLOEALATYGNSPAAMTTRYDGGFGMGKTYPTPLTRGVDCYALATYVDMHF 415
 338 GERVAEVSQGEAVLYGHTPRAGMOTKYIDVGMGSGVTHLAPGDCETTFIDAFH 397
 1 GERVAEVSQGEAVLYGHTPRAGMOTKYIDVGMGSGVTHLAPGDCETTFIDAFH 397
 416 LLESQAPKTIIRDAFCVFEONOGIPLRRHSDLYS---HIFGGLAEVLVLRNSSTLLND 472
 1 LLESQAPKTIIRDAFCVFEONOGIPLRRHSDLYS---HIFGGLAEVLVLRNSSTLLND 472
 398 YYDSQDPVHYPHALCLFEFPTGVPRLRRHFSNFGGFNFAGLKGVLVLRSTYVNYD 457
 1 YYDSQDPVHYPHALCLFEFPTGVPRLRRHFSNFGGFNFAGLKGVLVLRSTYVNYD 457
 473 YVWDYFHPSGAIEIRFYATGYSSAFLRGATGKYGNOVSEHNLGTVHHTSAFXYDDV 532
 1 YVWDYFHPSGAIEIRFYATGYSSAFLRGATGKYGNOVSEHNLGTVHHTSAFXYDDV 532
 458 YLMDYFHTSYNGVEAKMHAATGYHATFYTPREGLRHGRTRLOTLHLGHLVLRVDMDV 517
 1 YLMDYFHTSYNGVEAKMHAATGYHATFYTPREGLRHGRTRLOTLHLGHLVLRVDMDV 517
 533 AGLENVMAEDVAVFVMAVPMSPERHQLQRLQVTRKLEMEBOAFLVGSATPRYLYASN 592
 1 AGLENVMAEDVAVFVMAVPMSPERHQLQRLQVTRKLEMEBOAFLVGSATPRYLYASN 592
 518 AGTKNSFOTLTMLKLENTLNPMSHSLVQPTLEOTYQSGHQAPFRGQGLPRYLLFESSP 577
 1 AGTKNSFOTLTMLKLENTLNPMSHSLVQPTLEOTYQSGHQAPFRGQGLPRYLLFESSP 577
 593 HSKKCHHPRGYRIQMLSPAGEPLPONSMAKRGFSWERYQLAVYQKKEEPSSSSVENQND 652
 1 HSKKCHHPRGYRIQMLSPAGEPLPONSMAKRGFSWERYQLAVYQKKEEPSSSSVENQND 652
 578 OKKCMGHRSYRLOIHSMAEQVLPFGWOEBRATVMAHYPLAVTKYRESESEYSSSLYQND 637
 1 OKKCMGHRSYRLOIHSMAEQVLPFGWOEBRATVMAHYPLAVTKYRESESEYSSSLYQND 637
 653 PMAPTYDSDFI--NNEFTIAGKDLVAVWTAGFLIIPHAEDIPNVTYVNGNGGFLLRPNNF 711
 1 PMAPTYDSDFI--NNEFTIAGKDLVAVWTAGFLIIPHAEDIPNVTYVNGNGGFLLRPNNF 711

DB 638 PMDPVVEEPEELKNNENIDEDLVAVWTAGFLIIPHSSEVPNATGNSVGLLRPNNF 697
 1 PMDPVVEEPEELKNNENIDEDLVAVWTAGFLIIPHSSEVPNATGNSVGLLRPNNF 697
 OY 712 DEDPSFYSADSIYFRGDQACAEVNPALCLPQAAACADLPFAFSGH 759
 1 DEDPSFYSADSIYFRGDQACAEVNPALCLPQAAACADLPFAFSGH 759
 DB 698 PEDPSLASRDYIV-WPOKGLNVRQ--RMIPEDRCIVS-PPFSYNG 741
 1 PEDPSLASRDYIV-WPOKGLNVRQ--RMIPEDRCIVS-PPFSYNG 741

RESULT 4
 A54053
 amine oxidase (copper-containing) (EC 1.4.3.6) ABP1 precursor - human
 N:Alternate names: amiloride-binding protein; diamine oxidase; histaminase
 C:Species: Homo sapiens (man)
 C:Date: 07-Jul-1995 #sequence, revision 07-Jul-1995 #text_change 18-Feb-2000
 C:Accession: A54053; A54410; I38708; A38276; S42495
 R:Chassande, O.; Renard, S.; Barbry, P.; Lazdunski, M.
 J. Biol. Chem. 269, 14484-14489, 1994
 A:Title: The human gene for diamine oxidase, an amiloride binding protein. Molecula
 A:Reference number: A54053; MUID:94237856; PMID:8182053
 A:Accession: A54053
 A:Molecule type: DNA
 A:Residues: 1-751 <CHAS>
 A:Cross-References: EMBL:X78212; NID:g463242; PIDN:CAA55046.1; PID:g463243
 R:Novotny, W.F.; Chassande, O.; Baker, M.; Lazdunski, M.; Barbry, P.
 J. Biol. Chem. 269, 9921-9925, 1994
 A:Title: Diamine oxidase is the amiloride-binding protein and is inhibited by amilo
 A:Reference number: A54410; MUID:94193685; PMID:8144586
 A:Accession: A54410
 A:Molecule type: protein
 A:Residues: 20-27, 'A', 29-39 <NOV>
 A:Cross-References: PIDN:AA830395.1; PID:g546213
 A:Experimental source: placenta
 A:Note: sequence extracted from NCBI backbone (NCBIP:146046)
 R:Zhang, X.; Kim, J.; McIntire, W.S.
 Biochem. Genet. 33, 261-268, 1995
 A:Title: CDNA sequences of variant forms of human placenta diamine oxidase.
 A:Reference number: I38708; MUID:96113540; PMID:8595053
 A:Accession: I38708
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-331, 'F', 333-751 <RES>
 A:Cross-References: EMBL:U11862; NID:g533535; PIDN:AA50270.1; PID:g533536
 R:Barbry, P.; Champe, M.; Chassande, O.; Munemitsu, S.; Champigny, G.; Lingueglia,
 Proc. Natl. Acad. Sci. U.S.A. 87, 7347-7351, 1990
 A:Title: Human kidney amiloride-binding protein: cDNA structure and functional expr
 A:Reference number: A38276; MUID:91017502; PMID:2217167
 A:Accession: A38276
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-265, 'K', 267-751 <RES>
 A:Cross-References: GB:M55602; GB:M56335; NID:g387655; PIDN:AA58358.1; PID:g177960
 A:Note: this sequence has been corrected in reference A54053
 C:Genetics:
 A:Gene: GDB:ABP1
 A:Cross-References: GDB:127105; OMIM:104610
 A:Map position: 7q31-7q32
 A:Introns: 524/1; 619/2; 663/3
 C:Function:
 A:Description: oxidation of putrescine and histamine
 C:Superfamily: amiloride-binding protein
 C:Keywords: homodimer; oxidoreductase; quinoprotein; topaquinine
 F:1-13/Domain: signal sequence #status predicted <Sig>
 F:461/Modified site: topaquinine (Tyr) #status predicted

Query Match 34.0%; Score 1394; DB 2; Length 751;
 Best Local Similarity 39.8%; Pred. No. 3e-102;
 Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;

OY 50 PMTHPGOSQLFADLSRELTAVMFLTQRLGPGVDAQAARSDNCVFSVELQLPKAA 109
 1 PMTHPGOSQLFADLSRELTAVMFLTQRLGPGVDAQAARSDNCVFSVELQLPKAA 109
 DB 23 PGLTPRKAGVFSLSNOELKAVHSFLMSKKEELQPSSTTKMKNFTLEMLPKKYHV 82
 1 PGLTPRKAGVFSLSNOELKAVHSFLMSKKEELQPSSTTKMKNFTLEMLPKKYHV 82
 OY 110 LAHLDSGPPAREALAVFEGRQPNVSELVYGPLPHPSYMRDVERHGSGPLPYHRR 169
 1 LAHLDSGPPAREALAVFEGRQPNVSELVYGPLPHPSYMRDVERHGSGPLPYHRR 169


```

Db      83 LRLDGERHPRVEARAVIFFGDDEHPNTEFAVGLPGCYRALS -PRPGYQSSWASR 141
QY      170 PVLFOEYLDIDOKIFNRELPOASGLHH-----CCFKHGRNRLVMTTARGQSGD 222
Db      142 PLSTASY-----ALVHTLOEAKRPLHOEFLMTTGFSDCHRCALFTDVARGVASG 196
QY      223 RMTFELLYNISGAGFFLHVGLLELVNKKALDPAKMTIOKVYOGRYDLSLAOLEAOF 282
Db      197 RSMWLLIOKVE--GFLNPTGLELLVDHGSTDAGHNAVEQVYNGKFTGSPBELARKYA 234
QY      283 AGLVNVVLLPDNGTGSWSLKSFPVP-----GPAP----PLQFYPGGPRFSVOC 327
Db      255 DEEDVVDVLEDELPGGKGKGDSTEEEPPLFSSHKRPGDPSPSHVSGRLVQPHGPRFLEG 314
QY      328 SVYASSLMTFSGLGAFSGPRIFDVPRGGERLYEISLQALAYGNSPAAATTKRYVDG 387
Db      315 NALVGGMSFARLRNSSGQVNLNHFGBERLAEVSDAVLGYGHPAGMOTYLDV 374
QY      388 GFGMGKTYTTLRGVDCPYLATYVDHFLLESQAPKTIKDAFCVFQONOGPLRRHNSDL 447
Db      375 GMLGSLVTHELAPGIDCPEPATFLDFHYYDADDPVHYPRALCLFEMPICVPLRRHNSN 434
QY      448 YS---HYEGGLAEVTVLVNMSSTLLNVDYVMDTVFHPGSAIEIRFYATGYISSAFLEGAT 504
Db      435 FKGGFNFYAGLKQVLYLRTSTYVNYDYIMDFIFYPNGVMEAKHATGYVHATFYTPREG 494
QY      505 GKGNQVSEHTLTGTVTHSAHFKVDLDVAGLEMMVAAEDVFFPMAPVPSBEHQLOQV 564
Db      495 LRGGTLHHTLGNHTLHVHFDVLDVAGTKSFCOTLQKLENTINPMSPRRHVVQPTL 554
QY      565 TRKLEMEBOAFVLGSAATPRYLYLASNHSNKKWHPRGVRIQMLSPAGEPLPONSSWARG 624
Db      555 EQOYQMEQOAFNRFRKRLPKYLLFTSPOENPGHKRSYLOLHSHADQVLPFGNQEQA 614
QY      625 FSWERYQLAVTORKEEPPSSSVFNONDPAPTVDSDFI-NNETIAGKDLVAVMTAGFL 683
Db      615 ITWARYPLAVTKRESELSSSYHQNDDPHPVVEFOEFLHNENIENEDLVAVMTVGL 674
QY      684 HPHADIPPTVTVGVGVGFFLRPNFDEDPSEFSYSDSTIFYGDDAGACENPLACL 743
Db      675 HHSSEDIPTATPGNSVGLFLRPFNFPEDPPLASRDYIV-WPRNGPNVQ--RWIP 731
QY      744 QAAACAPDLPASFSG 759
Db      732 EDRDCSMP-PPFSYNG 746

RESULT 5
B41836
amine oxidase (flavin-containing) (EC 1.4.3.4) precursor - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 02-Sep-2000
C:Accession: B41836
R:Sugino, H.; Sasaki, M.; Azakami, H.; Yamashita, M.; Murooka, Y.
J. Bacteriol. 174, 2485-2492, 1992
A:Title: A monamine-regulated Klebsiella aerogenes operon containing the monamine oxid
A:Reference number: A41836; MUID:92210491; PMID:155668
A:Accession: B41836
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-755 <SUG>
A:Cross-references: GB:D10208; NID:9216721; PIDN:BA01060.1; PID:d1001529; PID:g216723
A:Superfamily: amine oxidase (copper-containing)
C:Keywords: oxidoreductase

Query Match      10.1%; Score 413; DB 2; Length 755;
Best Local Similarity 27.5%; Pred. No. 2,1e-24;
Matches 150; Conservative 75; Mismatches 213; Indels 108; Gaps 21;

QY      246 ELLVNHKALDPAKMTIOK---FYGGRYVDSLAQLEAFAGLVNVVLLPDNGTGSWS- 301
Db      237 EVLKKHGTDPGKVVVTTPLTVGFFDGK--DGLQO-----DARLLKVVSYLDTGDGNVMAH 289

```

```

QY      302 -----LKS-----PVPPGAP-----PLQF-YPOGPRFSVQ 326
Db      290 PLENVAVVDLEAKKLIKIEGVLIPVMEPRPYDGRDNAPVAVPLETTEBEGKAYTIT 349
QY      327 GSRVASSLMTFSGLGAFSGPRIFDVRFQG---ERLYEISLQALAYGNSPAAATTT 382
Db      350 GDTHHQNMDPHLRNLSRVGPILSTYTYNDNGTKRQVMEGSLGMIIVYGDPDVGWYK 409
QY      383 RYVDG-FGMGKTYTTLRGVDCPYLATYVDHFLLESQAPKTIKDAFCVFQONOGPLR 441
Db      410 AYLDSDGYGKGLTSPIVGKDAFSNAVLIDETIADYTKPPTIPGAVAIFE----- 461
QY      442 RHNSDLYSHYFG-----LAETVLVRSSTLLNVDYVMDTVFHPGSAIEIRYATGYS 496
Db      462 RYAGPEYKHEMKRPVNSERRELVYRWISTVGNITYIFDMVPHDNGTIGIDAGATIGA 521
QY      497 SAELEFAT-----GKYNQVSEHTLTGTVTHSAHFKVDLDVAGLEMMVAAEDVFEV 547
Db      522 VKGVLAKTMHDPKABDTRGTLIDHNIVGTTHQIYNFRLLDLDVGENNTLVAADEPVK 581
QY      548 PMAVPSPEHQLORLQVTKRLEMEBOAFVLGSAATPRYLYLASNHS--NKNHPRGCLR 605
Db      582 PNT---AGPRFTSTMQVNYDTIDSEKAAQKEDPGTIR--LLSNHSKERNKGNPVSY-- 633
QY      606 QMISFAG--EPLPONSMA-----RGFSWERYQLAVTORKEE-----PSSSVFNG 650
Db      634 QIIPYAGGHPATGAKFAPDEWYHRLSPMKQLWYTRHPFERPECKYPRRSADHG 693
QY      651 NDEMAPTVDSDFINNETIAGKDLVAVMTAGFLHPHADIPPTVTVGVGVGFFLRPN 710
Db      694 LQGYAK-----DDESLTHNDVVMITTGTTHVAREEMPIPT--EWALLALKPMWF 743
QY      711 FDEDDPS 716
Db      744 FDEPT 749

RESULT 6
E64889
amine oxidase (copper-containing) (EC 1.4.3.6) tyra precursor - Escherichia coli (str
N:Alternate names: monamine oxidase; tyraminase; tyramine oxidase
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: E64889; I40923; S65442
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E64889
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-757 <BLAT>
A:Cross-references: GB:AE000235; GB:U00096; NID:q1787643; PIDN:AC74468.1; PID:g17876
A:Experimental source: strain K-12; substrain MG1655
R:Azakami, H.; Yamashita, M.; Roh, J.
J. Ferment. Bioeng. 77, 315-319, 1994
A:Title: Nucleotide sequence of the gene for monamine oxidase (maoA) from Escherich
A:Reference number: I40923
A:Accession: I40923
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-247,'E',249-257,'VT',260-276,'I',277-287,289,'I',291-455,'P',457-658,'D
A:Cross-references: GB:ID23670; NID:q474301; PIDN:BA04900.1; PID:g809499
A:Experimental source: strain K-12; substrain W3110
R:Steinbach, V.; Benen, J.A.E.; Bader, R.; Postma, P.W.; de Vries, S.; Duine, J.A.
Eur. J. Biochem. 237, 584-591, 1996
A:Title: Cloning of the maoA gene that encodes aromatic amine oxidase of Escherichia
A:Reference number: S65442; MUID:96235221; PMID:8647101
A:Accession: S65442
A:Molecule type: protein
A:Residues: 41-48 <STE>

```

A:Experimental source: strain W3350

C:Genetics:

A:Gene: tynA: ma0A

C:Complex: homodimer

C:Function:

A:Description: catalyzes oxidation of phenylethylamine and water to phenylacetaldehyde,

A:Pathway: amino acid metabolism; phenylethylamine catabolism

A:Note: 2,4,5-trihydroxyphenylalanine quinone (topaquinone) cofactor; copper cofactor; h

C:Superfamily: amine oxidase (copper-containing)

C:Keywords: copper; oxidoreductase; quinoprotein; topaquinone

F:1-33/Domain: signal sequence #status predicted <Sig>

F:34-757/Product: amine oxidase (copper-containing) #status predicted <Mat>

Query Match 9.7%: Score 397.5; DB 2; Length 757;

Best Local Similarity 27.5%: Pred. No. 3,6e-23;

Matches 142; Conservative 74; Mismatches 211; Indels 89; Gaps 20;

```

QY 265 FYGGRYDLSLAQLEAFEGALVNVVLIIPDNGTGSMS-----LKS 304
DB 259 YEDGK--DGLKQ-----DRLKLVISYLDVGDENVNHPLENIIVAVVLEQKKIVKIEEG 311
QY 305 PVPFPGP-----AP---PLOFY--PQGRFVSQGSRYVASLMTFSFGLGAFSGPR 348
DB 312 PVVPVPMATARPFGDRQVAVAPKPMQIIEPEGKNTITGDMIMRMMDFLSNNSRYGPM 371
QY 349 IDYPRQG-----ERLYEISLDEALAIYGNNSAAMTKRYVGG--FGMKYTPPLRGVD 403
DB 372 ISTVTVNDNGTKRKRYEGSLGMIIVYGDPIGWYFKAYLDGSDYGMGTLPSPARKGD 431
QY 404 CPYLATVYDWHFLESQAKTIRDAFCVEQONGRLPLRHHNSDLYSHYFEG-----LAET 458
DB 432 AFSNAYLLMETIADYGVMEIRATAVE-----KAGREYKHQEGQGNVSTERR 483
QY 459 VLVYRSMSTLLNDYVDVTFVHPSGAIEIRFVATGY-----ISSAPLFGATGK----YGN 509
DB 484 ELVVRWISTVGNDYIFDMIFHENGCTIGIDAGATIEAVGVAKAKTMHDETAKDPTRYGT 543
QY 510 QVSEHNLGVTHSHAFKVDLDVAGLENNWMAADMFVPAVAVWSPRHQRLQVTRKLL 569
DB 544 LIDHNLTGVTHOITVFRLLDLDVGENSNLSVAMDPAVKPPT---AGSPRTSTVOVNOYNI 600
QY 570 EMEEOAALFVGSATPRYLYLASNHNKMGHPRGRYIOMLSFAG--EPLPONSSMA----- 622
DB 601 GNEQDAQAQFDPGRTIR--LSNPKENRMGNPSY--QIITVAGCTHNVAKGAPAPDEMI 657
QY 623 -RGFSNERYQLAVTQRKEEPPSSSVF--NQNDMPARTVDESDFINNETIAGKDLVAVNTA 660
DB 658 YHRLSFNDKQLWVTRYHNGRERPEKGYPNRSTHDTGLGQYSK--DNESLDNTDAVVMYMT 715
QY 681 GFLIHIAEDIPNTVTVGVNGVGFLLRPYNFDEDDPS 716
DB 716 GTTHVARAEWPIPT--EWVHTLLKPMWFDEPT 749

```

RESULT 7

A56102

amine oxidase (copper-containing) (EC 1.4.3.6) - Arthrobacter globiformis

C:Species: Arthrobacter globiformis

C:Date: 13-Apr-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000

C:Accession: A56102

R:Choi, Y.H.; Matsuzaki, R.; Fukui, T.; Shimizu, E.; Yorifuji, T.; Sato, H.; Ozaki, Y.;

J. Biol. Chem. 270, 4712-4720, 1995

A:Title: Copper/topa quinone-containing histamine oxidase from Arthrobacter globiformis.

A:Reference number: A56102; MID:95181469; PMID:7876243

A:Accession: A56102

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-684 <CHO>

A:Cross-references: GB:D38508; NID:9994746; PIDN:BA07517.1; PID:9994747

A:Note: authors translated the codon TGC for residue 239 as Phe

C:Superfamily: amine oxidase (copper-containing)

C:Keywords: oxidoreductase

Query Match 9.2%: Score 377; DB 2; Length 684;

Best Local Similarity 24.1%: Pred. No. 1.3e-21;

Matches 182; Conservative 107; Mismatches 343; Indels 124; Gaps 28;

```

QY 39 PHCPVSFSAQPMPT--HPGQSOLFADLSREELAVNRFLQRLGCVLDAAQAPSDNVC 96
DB 7 PSTPLVQDPPVPTLVHAAQHPLEQLSABEIEARILAE--AGLVG----- 52
QY 97 FSVELDLPKAAALAHNDGSP-----PAREALAIYFGRQPQ-----PNVSE 140
DB 53 -----ESTFATVLGLIEPKTRQGVGTGAARLYRMLMDAQSRLDVRSLATG 103
QY 141 LVVGPLPHPSYMDVTVERRHGGLPYHRRPVLYFOEYLDIDOMTFNRELPAQSLHHCCF 200
DB 104 LVVD-----RRELNPADG-----QLPVLLEEGLIEDIL--SEDPQMNAL----- 143
QY 201 YKHGRNLTMTAP-----RGLOGDRATWFGLYYNSGAFELHHVGLLELVNHA 253
DB 144 -TARGLTPAQCVRVAPUSAGYFEGNEGKRL-----LRGLGFRODP-----ADHPM 189
QY 254 LDPARTIQKVFYQGRYDLSLAQLEAFEGALVNVVLIIPD--NGTGSWSLKSVPFPGAP 312
DB 190 AHPIDGLVAVDVENRRVNL-----IDDGVPVY--PENVNGYTPRAIRGELRTDLP 240
QY 313 PLOFYQGRFVSQGSRYVASLMTFSFGLGAFSGPRFEDV--FQGR--LVYEISLQEA 368
DB 241 EIMQEGPSFTLEGHNLISAGMDLRVGFARGLVLIQHNHSHKRRRVIRHASTSEM 300
QY 369 LAIYGGNSPAAMTTRYVDG--FGMKYTPPLRGVDCPYLATVYDWHFLESQAKTIRD 427
DB 301 VVRYGDSPTYSQNGNFTDSEYLVGRANSLRLGCDLGITYSPPVADDFGPRITEN 360
QY 428 AFCVFEQNGRLPLRHHNSDLYSHYFGLAETVLVYRSMSTLLNDYVDVTFVHPSGAIEI 487
DB 361 GICIHEDDAGILWK--HTDEMAGSDEVRNRRRLVVSFTTGVGNVYGFYWTLYLDGTIEF 418
QY 488 RFATGVISSAFLEGAAGKGNQVSEHTLCTVTHSHAFKVDLDVAGLENNWMAEDVVF 547
DB 419 EAKATGIVFTAAALPDKDYAAASELAPGLGARYQHNLRSALDMMIDDDARVEEDLVRL 478
QY 548 PMAVPMSPHQRLQRLYTRKLEMEQAAFLVGSATPRYLYLASNHS--NKGHPRGRIQ 606
DB 479 PKG--PGPHG--NAFTQKRLRLARESEAVYADACAKRWNIHNSPDSLNLGHVGTIY 535
QY 607 MLSFAGEP---LPONSSMARGFSWERYQLAVTQRKEEPPSSSVFNQNDMPARTVDESD 663
DB 536 P---EGNPTLAMADSSIASRAAFARHHLWVTRHAEELYAAGDFVQHPGAVLP--AYV 591
QY 664 INNETIAGKDLVAVNTVGFHLIHPAEDIPNTVTVGVNGVGFLLRPYNFDEDDPSYSDSI 723
DB 592 AODRDIIDGODLVVHNSFGTLHFRPEDWP--IMPVDYDTGLKHKHGFEDENPLVPSSA 649
QY 724 YFRGDDAGACEVNPFLACLPOAAACAPDLPAFSHG 759
DB 650 -----AGHCGTG-----SERENAPGTRAVGHSG 673

```

RESULT 8

G90330

amine oxidase (copper-containing) (tynA) [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001

C:Accession: G90330

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.V.;

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redd

arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Coost, J.

submitted to Genbank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90330

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-660 <KUR>
 A:Cross-references: GB:AE006641; NID:g13814942; PIDN:AAK11902.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: tyna
 C:Superfamily: amine oxidase (copper-containing)

Query Match 9.0%; Score 368.5; DB 2; Length 660;
 Best Local Similarity 24.4%; Pred. No. 5, 8e-21;
 Matches 175; Conservative 103; Mismatches 303; Indels 137; Gaps 28;

```

OY 63 LSREELAVNHFLLTQRLGPGLVDAQAAPSDNCVSEVLQPPKAAALAHLDKSGSPPAR 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 17 LNEEIKRSSEVLRKQLN---LDAKVK-----FFSEVLEKPKKOEYLEMNNKNIKIER 68
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 123 EALALVFEGQPPNPVSELVVGPLPHPSYMDYVERHNGRLPHRRRVLFQELDIDQM 182
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 69 ESL-VKYNFQDR-KVYEALTI-----SLDNVVKKEIRSDAAY--PRTLDEGECKKA 118
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 183 IFNRELPOASGLHHCFCFYKRGKRNLYMTAPRGLOSGDRATWGLYINISGAGFFLH 242
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 119 VRNDRKQGEA-----LTKRGILINDLNT---LWVDCMARG-----H 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 243 VGEELVNHKALDPARNTIQKRYQRYDSLAOLEAOFEGALVNVLLIPDNGTGSWSL 302
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 152 VDEELRGRRAVIG-YMN-VKKDIEDNGYGRPVHGLMPVVDLKKMEVIRIDHGT----- 203
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 303 KSPVPPGPA-----PPLQF-YPOGRPSVQGSRAVSSLTWTSFGLGAFSG 346
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 204 -SPLPDADANTYPEKLRKIFGDDLPLEIRQLPSSSIKINMEISWYMRRLRIGTTPREG 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 347 PRIDVFRQGER-----LVYEISLOEALATYGNSPAAMTTRRYDGG-FGKGYTPTLR 400
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 263 LVITDVKRIDENNRERMLYASVVDLMPYGDSPFNKKMVLADAGYGLGNFIVPLSH 322
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 401 G-----VDC-PYLATVYDWHFLESQAQKTRDAFCVEQNOGLPLRRHSHDLYSHYG 453
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 323 GNYDLYNDCGEVYIHDVTRVSSNGTPIKIKAKICVHEDEFGVLMR--HTDLRS----- 376
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 454 GLAET-----VLVYRSMSTLLNYDVMPTVPHPSAIEIRFATGYSIAFL--GATGY 507
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 377 GKSEVRNRRLRYLVSWMTLATYDGFYEFYQDSIEFLVLTGLINDSISEKDPTRYK 436
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 508 GNOVSEHTLGTVHTSAHFYDLVAGLENNVMAEDVFMVAVWSPSEHQLQRLQVTRK 567
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 437 GTRVTPYVAIHIFHFRIRIRINIVDGLRNTY-----EVLNRGEPITREK 481
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 568 -----LLEMEQAALVGSATPRYLTLAS-NHSHKMGHPRGYRIOMLSFAGEP 614
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 482 NPVGNAFEAEENLEENADARRHVPQGRYKIVNIQKKMYLGLPVAYRLPGHNVLPR 541
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 615 LPONSMARGSWERYOLAVYQKKEEPPSSSVFNQNDPMAPTVDFSD-----FINNETI 669
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 542 LPDSDYARRKCAITNYHLMTVPYNEERYASGDY-----PYLRANDGLPKYILKKRSI 594
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 670 AGKDLVAVMTAGFLIHPAEDIPNTVTYGVNGVGFELRPYNFEDEDPSTYSADSIYFG 727
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 595 VDEDLVMTITGVEHVYRIEDMP--VMPVEMAGFRLIPDGFEDKMPITYLRLQRLKING 650
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 9

A48646
 amine oxidase (copper-containing) (EC 1.4.3.6) - *Arthrobacter* sp. (strain P1)
 C:Species: *Arthrobacter* sp.
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999
 C:Accession: A48646
 R:Zhang, X.; Fuller, J.H.; McIntire, W.S.
 J. Bacteriol. 175, 5617-5627, 1993
 A:Title: Cloning, sequencing, expression, and regulation of the structural gene for the
 hydroxoph.
 A:Reference number: A48646; MUID:93374858; PMID:8366046
 A:Accession: A48646
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA

A:Residues: 1-648 <ZHA>
 A:Cross-references: GB:U12990; NID:g289152; PIDN:AAA22074.1; PID:g289153
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: oxidoreductase

Query Match 8.9%; Score 363; DB 2; Length 648;
 Best Local Similarity 24.0%; Pred. No. 1, 6e-20;
 Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

```

OY 55 GOSQLFADLSREELAVNHFLLTQRLGPGLVDAQAAPSDNCVSEVLQPPKAAALAHLD 114
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12 GVSHPLDLSHVEIARAVAIKE--GPAAESFR-----FISVELREPSKDL----- 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 115 RGSPPRAREALATVFFGQPPNPVSELVVGPLPHPSYMDYVERHNGRLPHRRRVLFQ 174
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 58 RAGVAVAREADAVLV--DRAQSRFEAYVD-----LEAGTVSKILLAEINIDPPEFLD 108
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 175 EYLDIDMIFNRELPOASGLHHCFCFYKRGKRNLYMTAPRGLOSGDRATWGLYINISG 231
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 109 EFACEDCNC--RKDPE-----VIALAKRGLINDLVCFEFPMSVGYF 148
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 232 NISGAGFFLHVGLLELVNKHALD-PARNTIQK--VFYQGRYDSLAOLEAOFEGALVNV 288
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 149 GEDNEGRL--MRALVFYRDEADSPYAHPIENFIYEFY-----DLNAG--KV 191
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 289 VLIDPNTGSGWSLSKSPVPPGPAAPLQFY-----PQGRPSVQGSRAVSS 333
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 192 VRLIEDQA-----IPVPSARGNTLPKRYVGEARTDLKPLNITQPEGASFVTGNHVTWA 244
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 334 LMTFSFGLAGSPRIEFVFRQGE---RLVYEISLOEALATYGNSPAAMTTRRYDGG 388
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 245 DMSFRVGTTPREGVLVHLQKFKDGVDRPYINRASLSMVPYGDTAVOAKKNAFDSGE 304
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 389 FGKGYTPTLRGVDCPYLATVYDWHFLESQAQKTRDAFCVEQNOGLPLRRHSHDLY 448
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 305 YNIGMANSLTLCDCLEIKYFQGHSDVSHGNPTIENALCMHEEDDSI-----LM 356
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 449 SH--YFGSLAET-----VLVYRSMSTLLNYDVMPTVPHPSAIEIRFATGYSIAFL 501
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 357 KHFDRREGTAETFRNSRKLVISFIATVANYETAFYWHLEFLDSDIEFLVATGSLTAGOLP 416
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 502 GATGKYGNQVSEHTL-GTVHTSAHFYDLVAGLENNVMAEDVFMVAVWSPSEHQL- 559
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 417 GEKNPYGGSLNNDGLYARVHQHMFVRMDFELDGVKKNVYEVDEMEX-----PEHNPT 468
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 560 -QRLQVTRKLLLEMEQAALVGSATPRYLTLASNHS-NHSHKMGHPRGYRIOMLSFAGEP 617
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 469 GTAFMAVDRLLETBOKAIRKTNEAKHREMKIANHESKNLVNEPVAYRL-----IPT 519
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 618 NSSMARGSWERYOLAVYQKKEEPPSSSVFNQNDPMAPTVD-----FSDFINNET- 668
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 520 NG-----IQLLA--RDDAYSKRAQFARKNLMTAYADRTERFAAGETPNQATGAD 567
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 669 -----IAGKDLVAVMTAGFLIHPAEDIPNTVTYGVNGVGFELRPYNFEDEDP 716
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 568 DGLHMTQKDRNIYDVTDLVWYTFEGMHVHVRLEDMP--VMPRONIGFMLEPHGFENQPT 625
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 717 FYSADSIYFRGDDQAGAC 734
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 626 INLPSTSTQTQGEADTC 643
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 10

T48139
 copper amine oxidase-like protein - *Arabidopsis thaliana*
 N:Alternate names: protein T4C9.130
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
 C:Accession: T48139
 R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dijkse, W.; Stiekema, W.; Bancroft, I.;
 submitted to the Protein Sequence Database, June 1999
 A:Reference number: Z24485
 A:Accession: T48139

A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-756 <BEV>
 A:Cross-references: EMBL:AL080318
 A:Experimental source: cultivar Columbia; BAC clone T4C9
 A:Genetics:
 A:Map position: 4
 A:introns: 460/3; 498/3; 686/2
 A:Note: T4C9.130
 C:Superfamily: amine oxidase (copper-containing)

Query Match 8.7%; Score 357.5; DB 2; Length 756;
 Best local similarity 22.2%; Pred. No. 5.3e-20;
 Matches 163; Conservative 100; Mismatches 280; Indels 191; Gaps 29;

QY 94 NCVEVELQPPKAAALAHLDGSPPPAREALAI-----VFGRQPPQPVSE 140
 DB 126 HALHTVLEPER-NLYRHKKNPLPRKASYIAKVGADTHVLTVDISTGRVDSEKSPV 184
 QY 141 LVVG-PLPHPSYRDVYVERHGSLPYHRRPVLFQEYLDIDOMIENRELPOASGLLHCC 199
 DB 185 RVSGYPMWTIEEMNDITV-----VPSNAD-----FNRTIIS----- 216
 QY 200 FYHNRGRNLVTMTAPRGLOSGDRATWFG-----LYNISGAGFLIHHV-GLE 246
 DB 217 ----RGNLTDVICFP--ISCG---WFGNKEENARYIKSQCFMTOCTPFNFRPIEGLT 266
 QY 247 LVNHRKALDPARMTIQKVFQGRYDLSAQLAEAGLVNVVLIIPDNGTGSMSLSKSPV 306
 DB 267 ILID--LDTKQ-----PAPLPQF-TPQGRFVSQGRVAS-SLMTSSEGLGAFSGP 347
 QY 307 PFG-----PAPLPQF-TPQGRFVSQGRVAS-SLMTSSEGLGAFSGP 347
 DB 289 PGSTNTDYRFQKLTATDKTRPLNPISIEOPRGSEVIEDHNLVKNAMERHLRPDPRAVY 348
 QY 348 RIDVR-----FOGERLVEISLOEALAYGGNSPAAMTTRYVDG-FGMGKTYTPPLTR 400
 DB 349 VIRRVKHDPDTHETRDVMTKGFVSELPVYMPDSAMYKTYMDAGEYFGLOAMPVLP 408
 QY 401 GVDCPYLATVYDWHFLESQAPKTIRDAFCVFQON-----QGLPLRRHNSDL 447
 DB 409 LNDCCPRMAAYMGVFAAAGSTPFVRENMCIFESYAGDGMHSESPITGIPREVR--- 465
 QY 448 YSHYFGGLAETVLYVRSMSLTLNIDYWDVTFHPSGAIETRFATGITSAPLEGATGKY 507
 DB 466 -----PKVTLVVRMAASVGNVDYIIDEFQTDGLIKAKVGLSGIL---MYKGTYYQN 514
 QY 508 GNOV-----SEHTIGTVHTHSAHFKYDVLVAGLENNVMAEDMVFVMAV 551
 DB 515 KNOVEDKDKGNEELHGLTLESENVIGYIHDHYTYFYLDLVDGPDN-----SFYKVALKR 569
 QY 552 PMSPEHOLQ--LQVTRKLLMEEOAFLVGSATP-RYLYLASNHSNKGHPRGYIOM 607
 DB 570 QETEPGSPRSKYLKAARNIAKTEKQDQIKLSTLDPSEFVINSKTKTRGNGPFGYVVP 629
 QY 608 LSTRAGEPLPONSMAKFSERYQLAVTQKKEEPPSSSVFNQNDPAPRYVD-FSDINN 666
 DB 630 RTTAASTLDDHDPQKGAFTNNQIWTVPYKSEOMAGGFTYQSHGDDTLAWASD--RD 687
 QY 667 ETAGKDLVAMVTAGFLIHAEDIPNTVTVGNQVGFELRPYVNFEDDPSFYSDSTYFR 726
 DB 688 RDLNKNDIVMYTLGFHHTICQEDFPIPTVSS--SFDLRFVNFEEKNPILSAAPN--- 742
 QY 727 GDQDAGACEVNPPLA 740
 DB 743 -EHDLPVCGVGSYS 755

RESULT 11
 G71412
 Probable amine oxidase - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Variety: Columbia

C>Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
 C:Accession: G71412
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.;
 P.; Medler, H.; Medler, E.; Wambutt, R.; Weizenecker, T.; Pohl, T.M.; Terry, N.;
 Avonagh, T.; Hempel, S.; Kotter, P.; Entlian, K.D.; Rieger, M.; Schaeffer, M.; Funk,
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Putgdon
 erhof, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.
 C.; Chalmers, N.
 A>Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis
 A:Reference number: A71400; MUID:98121113; PMID:9461215
 A:Accession: G71412
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-650 <BEV>
 A:Cross-references: GB:297337; NID:92244829; PID:92244851
 C:Genetics:
 A:Map position: 4COP9-4G3845
 C:Superfamily: amine oxidase (copper-containing)

Query Match 8.6%; Score 353.5; DB 2; Length 650;
 Best local similarity 23.9%; Pred. No. 8.9e-20;
 Matches 161; Conservative 96; Mismatches 278; Indels 139; Gaps 29;

QY 99 VELQLPKAAALAHL-----DRGSPPARRELAIVFGRQPPQPVSELYVGPLPHPSYMRD 154
 DB 56 LDLEPPNKSHTVQLSPNPKRPPRRSRFVVRAAGQ---TYELI-----D 101
 QY 155 VTEVER-----HGQ-----LPYHRRPVLFQEYLDIDOMIENREL-POASGLLHCC 199
 DB 102 LITSKLAASRLTYGHGPFSTFIELFKASLPLTPPFKSLIDRLSINISEVCIPRTVG 161
 QY 200 FYHNRGRNLVTMTAPRGLOSGDRATWFGLYVINSAGFGLIHHV-GLELLVNRKALDPAR 258
 DB 162 WYGE-----TTTRELKA-----SCFYRDSVNVFTRPIEGITVITD--VD--- 200
 QY 259 WTIQKVFQGRYDLSAQLAEAGLVNVVLIIPDNGTGSMSLSKSPVPGAPLPQFY- 317
 DB 201 -SMQVIRKYSDFRKP-----IPDK-EGNDFRKH-----RPFDFC 234
 QY 318 -PGGRFVSQGRVASSLMTFSFGLGAFSGPRIFDV-----RFOGERLVEISLOEA 368
 DB 235 NVSDTGKILGNKYNKAMKMFHGFTRAGVTISTASVLDPRKRRF-RVMRGNHSET 292
 QY 369 LAIYGNSPAAMTTRYVD-GFGMGKTYTPPLTRGVDCPYLATVYDWHFLESQAPKTIR 427
 DB 293 FVRYMDPTVEMYRTFMDIGFGRSAVNLQPLDPCPQNAAFLDGHNAGPDGTAQKMTN 352
 QY 428 ACFVFEQNOGLPLRRHNSDL---SHYFGLAETVLYVRSMSLTLNIDYWDVTFHPSGA 484
 DB 353 VMCVFERN-GYGASFRTETINPQOVITSEAEISLVAVATLGNVDYVDVEFKNGA 411
 QY 485 IEI-----RYATGYISSAFEGATGYGNQVSEHTIGTVTHSAHFKYDVLVAGLE 536
 DB 412 IRVGVDLTGVEAKATSYTNDQ--TENVYGLTVAANTATVANHDLTYLLDLDVQNG 469
 QY 537 NMYAEDMVFVMA-VMSPEHOLQVTRKLLMEEOAFLVGSATPRLYLASHNSN 595
 DB 470 NSLVKAKLKTVRVTEVKNKTSRSKSYWTYKETAETADGRVRLGSDPVELLTPNPKT 529
 QY 596 KMGHPRGYRIOMLSFAGEPLP--QNSSMARGFSEMEKQLAVYQKKEEPPSSSVFNQND 652
 DB 530 KIGNTVGYRL-----IPEHLQATSLTDDDPYELRAGYT---KYPVWVATYDRSE 576
 QY 653 PMAPTVDFSD-----FINNETIAGDLVAMVTAGFLIHAEDIPNTVTVGNV 701
 DB 577 RMAGGF-YDSRSKGGDDSLAWSSNNREIKNDIYMTNVGFHPIYQEDDFPYMPTLNG-- 633
 QY 702 GFELRPYNFDEDP 715
 DB 634 GFTLRPSNFDDNDP 647

RESULT 12

JC2139
Phenylethylamine oxidase (EC 1.4.3.-) - *Arthrobacter globiformis*
C:Species: *Arthrobacter globiformis*
C:Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 28-May-1999
C:Accession: JC2139; PC2070
R:Tanizawa, K.; Matsuzaki, R.; Shimizu, E.; Yorifuji, T.; Fukui, T.
Biochem. Biophys. Res. Commun. 199, 1096-1102, 1994
A:Title: Cloning and sequencing of phenylethylamine oxidase from *Arthrobacter globiformis*
A:Reference number: JC2139; MUID:94197690; PMID:8147851
A:Accession: JC2139
A:Molecule type: DNA
A:Residues: 1-638 <TAN>
A:Cross-references: GB:003517; NID:9451488; PIDN:AAA1814.1; PID:9451489
A:Accession: PC2070
A:Molecule type: Protein
A:Residues: 318-38-52,132-142,185-209,219-235,402-419,474-497,504-537,540-559,611-625
C:Comment: This enzyme catalyzes the oxidative deamination of various biogenic primary
C:Superfamily: amine oxidase (copper-containing)
C:Keywords: copper binding; oxidoreductase; quinoprotein; topaquinnone
F:355,431,433/Binding site: copper (His) #status: predicted
F:382/Modified site: topaquinnone (Tyr) #status: predicted

Query Match

Best Local Similarity 8.38; Score 340; DB 2; Length 638;
Best Local Similarity 23.78; Pred. No. 1e-18;

Matches 157; Conservative 100; Mismatches 274; Indels 132; Gaps 28;

QY 106 KAALAHLD--RGSPPARREALATVFF-----GRQOPRVSELVGRLHPSPMDVYER 159
DB 38 RIATLVGLDPAAGAGSEDEDRFRFTHDSGARQ-----EVTYSV 79
QY 160 HGGRPL-----PYHRRPLFOEYLDIDQIFN-----RELPOAS 192
DB 80 TNGVIVASVELDTAATGELPLVEEFVEVQDLATDERWKLALARNLDKVRARPLSA 139
QY 193 GLIHNCCEYKRRKRLVMTTAPRGL-----OSGRATWFGLYNISGAGFLHVGLELL 248
DB 140 GVEFYA---EERGRRL-----RGLAVQDFPEKSAN---AHAPDGLVAAYDVVSKETV 187
QY 249 VNRKALDPARTIQRVYOGRYDLSLAQLEAFAGLVNVLLIPDNGTGSMSKSPVP 308
DB 188 ---RVIDGVPRPA---ENGVYTD-----ELTGRLT 215
QY 309 GRAPPLQFYPOGPRPSVO--GSRVASSLMTFSFGAGSPRIFDVRFQ--GERL--VYEI 363
DB 216 TOKRISTOPGPGSPYTVGGNHIEKMSLDVGEVREGVVLHNAIFRDGRLRIINRA 275
QY 364 SLOEALAYGNSPAAATTRVVD--GGFGMKYTPRLTRGVCC---PLYATYVMHFLLE 418
DB 276 SIAEVVPRYGDSPRIKSNQNTFDGEYLVQYANSLELGCCLGDTITLSPVSDAF--- 332
QY 419 SQAEKTRDAFCVEQONGLPLRRHNSDLYSHFGLAETVLVVRSMSTLLNDYVMDTV 478
DB 333 -GNREIRNGICMHEEDMGLAK--HSDLMGINTTRNRKRMVLSFTTIGNIDYGFYVY 389
QY 479 FHPGSAIEIRYATGYI--SSAFLEFGATGKYGNQVSEHTLGTVHTSHAHKYVDLVALEN 537
DB 390 LYLDTITFEAKATGVVFTSAFPEGGSUNI--SQLAPGLGAFHQHIFGARLMDAIDGFTN 448
QY 538 VYMAEDVYFVMAVPMSEHOLQRYTRKLLMEQALFVGSATPRYLASHNS--NK 596
DB 449 RVEEDVVRQTMCGNERGNAFSR---KRYLTRESEVREADATGTWTWISNPKSR 505
QY 597 WGRHGRYRI---QMLSEFAGEPLPONSSMARGFSEWERYOLATQREKPPSSSVF--NOND 652
DB 506 LNEVVGKRLAHNQPILLADP---GSSIARRAATFKLDMYTRAADDRRYTPGDVNVNHS 562
QY 653 PMATTVDESDFINNETJAGKDLVAMVTAAGLHHPAEDIPNTVTVGNCVGFLLRPYNFED 712
DB 563 GGAGLPSY--TAODRDIDGDIVVMHTEFGLTHFRVEDMP--IMEVDIVGKRLREGFFD 618
QY 713 EDP 715

DB 619 RSP 621

RESULT 13

JC7251
amine oxidase (copper-containing) (EC 1.4.3.6) - garden pea
C:Species: *Pisum sativum* (garden pea)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: JC7251
R:Koyanagi, T.; Matsumura, K.; Kuroda, S.; Tanizawa, K.
BioSci. Biotechnol. Biochem. 64, 717-722, 2000
A:Title: Molecular cloning and heterologous expression of pea seedling copper amine o
A:Reference number: JC7251
A:Accession: JC7251
A:Molecule type: mRNA
A:Residues: 1-674 <KOY>
A:Cross-references: DDBJ:AB026253
A:Experimental source: seed
C:Comment: This enzyme, a homodimer, containing one Cu2+ ion and one 2,4,5-trihydroxy
diamine substrate and catalyzes the oxidative deamination of mono-, di-, polyamines t
C:Superfamily: amine oxidase (copper-containing); oxidoreductase
C:Keywords: copper; copper binding; homodimer; oxidoreductase

Query Match

Best Local Similarity 8.18; Score 332; DB 2; Length 674;
Best Local Similarity 23.08; Pred. No. 4.7e-18;

Matches 182; Conservative 105; Mismatches 293; Indels 212; Gaps 40;

QY 1 MNQTIIVLLAVITTTALVLCVLRGSGGEPSSQLPCHPSVSPSAQPTWHPGQSOLF 60
DB 1 MASTTKIAIFSVTITISFHAVV-----SVPLHYV--QHP-----L 35
QY 61 ADLSREELTAVRFLTORLGP-----GLVDAAQARPSDNCVSVLEQLPCKAAA 109
DB 36 DPLTKEEFLAV-QITVQKKYISKKLAFHYITGLDD-----PEKDYLRYE----- 80
QY 110 LAHLDGSPPARREALATVFFGGRQOPRVSELVGRLHPSPMDVYERHGRPLPYHRR 169
DB 81 -THPLTVSIP--KSFVVAIINSQTHEILDLRI-----RSIVSDNINNGGFP-- 127
QY 170 PVLPEYDIDQMFNRLPQASGLHHCFF---YKHGRNLYMTTAPRGLQSGDRATW 226
DB 128 -----LSVDSQSLATLP-----LKTPTIDSKYKKGKLNLSIYC-----SFTMGW 169
QY 227 FGLYYNI-----SGAGFLAHV-GLELVNHRKALDPARTIQRVYOGRYDLSL 274
DB 170 FGEENKVTATVRLDCMKKSTVINIVRPITGITVAD--LD-----LKKIV---EYHNR- 217
QY 275 AOLEAFAGLVNVYLLPD--NGTGSMSLKSPPVPRGAPR--LOFTYPOGPRFVQSGRYA 331
DB 218 -----DIEAVPTAENTEVQVSKSP--PFGPKQHSITSHQPGPGFOIEGHSVS 264
QY 332 SLMWTFSPGLAFSG-----PRIPDV--RFOGRLVYEHSLOEALAYGNSPAAATTRYV 385
DB 265 WANKFTHIGFDVRAGIVYSLASTIDLEKHSRRVLYKGIYISELVYPDPTPEEYFKTFE 324
QY 386 DGG-FGMKQYTPRLRGVDCPYLATYVDMHFLLESQAERTIDAFCVEQONGLPLRRHH 444
DB 325 DSGEGFGLSVSLPLPNDCPRHQFIDTYIHSANGPTILKKNACVREQ-----YG 376
QY 445 SDLSHYTGCL-----AETVLVVRSMSTLLNDYVMDTVFHPGSAIEIRYATGYI 495
DB 377 NIMWHTEGNGPNESIEESRTEVNLIYRTIVTVGVYDVIWMEFAASISIRPAIALSGIL 436
QY 496 SSAFLFGATGK-----GNOVSEHTLGTVHTSHAHFVVDLVALEMVMAEDVVF 546
DB 437 E---IKGTNINHKDEIKDLHGLKVSANSISGIYHDHFTIYLLDIDIDTHNSFEKTSIKT 493
QY 547 VPM-----AVPSPHQRLQVTRKLLMEQALFVGSATPRYLASHNSNKNKGH 599
DB 494 VRKDGSSKRSRYMTTEQTAK-----TESDAKITIGLAPAEVLVVPNITAYAGN 544
QY 600 PRGRIQMLSFAGEPL-----PONSSMARGFSEWERYOLATQREKPPSSSVFNONDP 653

Db 545 EVGYRL-ITAIHAHPLEDDYPO-----IRG-AFTYNNWVT-----AYNRTEK 587

Qy 654 WAP--TVDES---DFI-----NNETIAGDLVAMWTAGFLHIPHADIPTVTYVNGVGF 703

Db 588 MAGGLYVDHSDRDDLAVMTKONREIVNKDIYMHVYGHVPAQDDFIMPLLL--STSF 645

Qy 704 FLRPYNEDEDP 715

Db 646 ELRPNTNEFERNP 657

RESULT 14

T47403

amine oxidase-like protein - Arabidopsis thaliana

N:Alternate names: protein F23N14.50

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000

C:Accession: T47403

R:Blocker, H.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quetler, F.; Salanoubat submitted to the Protein Sequence database, March 2000

A:Reference number: 224465

A:Accession: T47403

A:Status: preliminary

A:Molecule type: DNA

A:Cross-references: EMBL:AL138638

A:Experimental source: cultivar Columbia; BAC clone F23N14

C:Genetics:

A:Map position: 3

A:Intons: 145/3; 303/2; 402/3; 440/3; 591/3; 624/2

A>Note: F23N14.50

C:Superfamily: amine oxidase (copper-containing)

Query Match 8.0%; Score 327; DB 2; Length 687;
 Best Local Similarity 21.6%; Pred. No. 1.2e-17;
 Matches 171; Conservative 113; Mismatches 314; Indels 194; Gaps 31;

Qy 7 LVLLIATVITFALVGVLCVLRGSDGGEPSQLPHCSVSQAQPMTHPGSOJFADLSRE 66

Db 9 LVLLIATVITFALVGVLCVLRGSDGGEPSQLPHCSVSQAQPMTHPGSOJFADLSRE 66

Qy 67 ELTAIVARFLTORLPGVLDAAQAPSDNCVSEVLEQLPRAALAHLDGSPAPRALA 126

Db 50 EIRKVTITLSGH-DPGEFGSS-----TIHMAIDEPKQKIRKKKGRLPRRAEIL 102

Qy 127 IYFGGQPOPNV---SELVYGPLRPHSY-----MRDVTVERHGGLPYHRRVLEQEYL 177

Db 103 AMNGESHVLTVDLKSGRVYSDLVNPTFGYPILTMKDIAV---SQVPY-----KSV 151

Qy 178 DITQMFNRELPOASGLHHCCEYKRGRLVMTTAARGLSGDRATWGLYNNISGAG 237

Db 152 EFRKSTEARQIP-FSGIICITPPAGWYGPD-----EGGRARYIKIQCFSKODTVN 199

Qy 238 FFLAHY-GLELLVNHKAALDPAWTIQKVFGHRYTSLAQLEAQLFAGLVNVLIPNGT 256

Db 200 FYKRPLEGILYVDMKLE-----TIKIYDNG- 226

Qy 297 GGSMSLSPVPPAPAPLAP-----YPOGPRFSVO-GSRVASSLMTF 337

Db 227 -----PVYPRKSTGEYRGLFENIYYMDRVNPMSEMGDPGSPQVEDYILVKMANMKF 280

Qy 338 SEGLAFSGPRIFDVNRQGERLYEISLOEALAI-YGNSPALM-----TTRYV 385

Db 281 HIKPDORAGMI-----SQAIVRDSKTGEARSVMYKGFASELFPVNMDEGEGYSKAYM 334

Qy 386 DGG-FGSGKTYTLRGVNDPILATYVDMHLEESQAPKIRDAFCVEFQNOGGLPLRRH 444

Db 335 DADEFGLGSSMPLVPLNDPARNAYTDGFASPEGIPILQPNKICLFERYAGDTSWRHS 394

Qy 445 SDLYSHYFGGL-----AETLVVYRSMSTLLNTDYVDYTFHPSGAIEIRFYATG----- 493

Db 395 EIL-----LPGVDIRSAKAKTIVARMAAGSVGNIDYIFDMQFQMDGIVRIVYVAASGMLMK 450

Qy 494 ---YISATFLFAGTGYGNVSEHTLGYTHSAHKVLDVAGLENNWMAEDMVFVPM 550

Db 451 GYATENEVDLGERKEDDSGLISENVIGVYHDEIFSHLMDIDGSAN---NSFYVHLE 506

Qy 551 VPMSPFHOLR---LOVTRKLLMEEQAAFLVGSATPRYLTLAS-HSKNKGHPRGYRIQ 606

Db 507 KQRLPGESRRKSYLKVKKYVAKTEKDAQIKMSLYPIEFHLVNPRLSLRGNPAGYKLV 566

Qy 607 MLSFAGEPLPONSSMARGEWMERYOLAVTQRKEEPPSSSVFNQNDPMA----- 655

Db 567 PGNAASLLDHDPPQMRGAFTNNOIWMTR-----YRSEOMAGLLMYQSGE 615

Qy 656 PYVD-FSDFINNETINGKLVAMVTAGFLHIPHADIPTVTYVNGVGFLLRYNFDD 714

Db 616 DTLQVMSD--RRRSINKDIYLVYTLGFFHVPQDEDFPVMPTIAS--SFLKRVNFEEN 671

Qy 715 P-----SFLYAD 721

Db 672 PVLGISPFPEKD 683

RESULT 15

C44239

amine oxidase (copper-containing) (EC 1.4.3.6) precursor - garden pea

C:Species: Pisum sativum (garden pea)

C:Date: 31-Dec-1993 #sequence_revision 05-Jan-1996 #text_change 17-Mar-2000

C:Accession: A57327; C44239

R:Tripping, A.J.; McPherson, M.J.

J. Biol. Chem. 270, 16939-16946, 1995

A:Title: Cloning and molecular analysis of the pea seedling copper amine oxidase.

A:Reference number: A57327; MUID:95348126; PMID:7622512

A:Accession: A57327

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-674 <TIP>

A:Cross-references: GB:J39931; NID:9685197; PID:9685198

R:Jones, S.M.; Palcic, M.M.; Scaman, C.H.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Biochemistry 31, 12147-12154, 1992

A:Title: Identification of topaquinine and its consensus sequence in copper amine c

A:Reference number: A44239; MUID:93090748; PMID:1457410

A:Accession: C44239

A:Molecule type: protein

A:Residues: 409-417, 'X', 419 <JAN>

A:Experimental source: seedling

C:Superfamily: amine oxidase (copper-containing)

C:Keywords: copper; glycoprotein; metalloprotein; oxidoreductase; quinoprotein; top

F:1-25/Domain: signal sequence #status predicted <Sig>

F:26-674/Product: amine oxidase (copper-containing) #status predicted <Mat>

F:156-389/Binding site: carbohydrate (asn) (covalent) #status predicted

F:382,467,469/Binding site: copper (His) #status predicted

F:412/Modified site: topaquinine (Tyr) #status experimental

Query Match 7.8%; Score 320; DB 2; Length 674;
 Best Local Similarity 25.8%; Pred. No. 4.3e-17;
 Matches 120; Conservative 66; Mismatches 189; Indels 90; Gaps 20;

Qy 301 SLKSPVPPGAPP--LQFYPOGRFVSQGSRAVSSLTFSFGAGFSG-----PRIFDV- 352

Db 233 SKQSP--PFGPKQKSLTSHQOGPOFOJNGSVAMNKKPHIGDVRAGIYISLASIYDLE 291

Qy 353 RFOGERLYVEISLOEALAIYGSNPAAMTRYVDG-FGSGKTYTLTRGVNDPILATYV 411

Db 292 KHKSRRLVLYGYISELFPQDPTEEFYKTFPDSGEGFGLSVSLIPNRDPPHQAQFI 351

Qy 412 DMHFLAESQAPKIRDAFCVEFQNOGGLPLRRHSDLSHFGGL-----AETLVV 462

Db 352 DTYVHSANGTPIILKNAICVFO-----YGNIMRHRHENGIPNSESIREVNLIV 403

Qy 463 RSMSTLNDYVDVTFHPSGAIEIRFYATGYISSAFLFAGTK-----YGNQVSE 513

Db 404 RTIVYGNVNDVDMWEKAGSITKPSIALSGILE---IKGTNKKHKEIKEDLHGKLVSA 460

Qy 514 HTLGYTHSAHKRVLDVAGLENNWMAEDMVFVPM-----AVPMSPFHOLRQLOVTR 566

```
Db 461 NSIGYHDFHYLYLDIDGTGTHNSFEKTSUKTVRIKDOSSKRKSWTETOTAK----- 515
OY 567 KLEMEEQAPLVGSATRIYLYLASNHSKNGCHPRGYRIQMLSTFAGEPL-----PQNSS 620
Db 516 ---TESDAKITIGLAPAPELVYVNPNIKTAVGNEVGYRL-IPAIPAHPILTEDDYPO--- 567
OY 621 MARGFSMERVOLAVTQRKEEPESSSVFNONDPMAP--TVDFS---DFI-----NNETIA 670
Db 568 -IRG-AFTNNYVWYT-----AYNRTEKMGGLYVDHSRGDDTLAVWTKONRETV 614
OY 671 GKDLVAVWTAGFLAIPHAEDIPNTVTGNGVGFFLRPNFDEDEP 715
Db 615 NKDIYMHVVGIIHVPAQOEDFPIPL--STSEFLRPTNFERNP 657
```

Search completed: May 20, 2003, 11:39:22
Job time : 33.4394 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 18.01 Seconds
(without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095
Sequence: 1 MNOKTILVLLITAVITIFAL.....QAAACAPDLPAFHSNGFSHN 763

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4095	100.0	763	1 AOC3_HUMAN	Q16853 homo sapien
2	3473	84.8	765	1 AOC3_MOUSE	O70423 mus musc
3	3346	81.7	762	1 AOC3_BOVIN	O46406 bos taur
4	3334	81.4	762	1 AOC3_BOVIN	O28437 bos taur
5	2612	63.8	756	1 AOC3_HUMAN	O71106 homo sap
6	1418	34.6	321	1 AOC3_RAT	O08590 rat
7	1400.5	34.2	746	1 ABP_RAT	P36633 rat
8	1394	34.0	751	1 ABP_HUMAN	P18801 homo sap
9	413	10.1	755	1 AMO_KLEAE	P49250 klebsiella
10	397.5	9.7	757	1 AMO_ECOLI	P46883 escherichia
11	377	9.2	683	1 AMO_ARTGO	O59118 arthroba
12	363	8.9	648	1 AMO1_ARTS1	O07121 arthroba
13	363	8.9	648	1 AMO2_ARTS1	O07123 arthroba
14	340	8.3	638	1 PAOX_ARTGO	P46881 arthroba
15	320	7.8	674	1 AMO_PEA	O43077 pisum sativ
16	311.5	7.6	666	1 AMO_LENCU	P49252 lens culin
17	310	7.6	671	1 AMO_PICAN	O12556 aspergillus
18	295	7.2	692	1 AMO_PICAN	P12807 pichia angu
19	111	2.7	1322	1 POR4_XYLLA	O96416 xyella fas
20	104.5	2.6	736	1 DVL2_MOUSE	O06083 mus muscu
21	103	2.5	2224	1 PAS_HUMAN	O14641 homo sap
22	103	2.5	2224	1 PAS_HUMAN	P12259 homo sap
23	102.5	2.5	591	1 COXN_BRAJA	P98000 bradyrhizob
24	101.5	2.5	1520	1 ACED_ECOLI	O46837 escherichia
25	101	2.5	435	1 KICH_MOUSE	O54804 mus muscu
26	100.5	2.5	597	1 DIVJ_MOUSE	O03228 caulobacter
27	100.5	2.5	1257	1 CAML_HUMAN	P33004 homo sap
28	98.5	2.4	881	1 NIAL_PHAVU	P39865 phascolus v
29	98.5	2.4	1217	1 S3B3_HUMAN	O13393 homo sap
30	98	2.4	2212	1 RRPL_EBOZM	O05318 ebola virus
31	97	2.4	716	1 DVL3_HUMAN	O92997 homo sap
32	97	2.4	716	1 DVL3_MOUSE	O61062 mus muscu
33	97	2.4	733	1 HEXA_BLADE	O17127 blaiberus di

ALIGNMENTS

RESULT 1	ID	ACOC3_HUMAN	STANDARD:	PRT:	763 AA.
AC	O16853:				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1) (HRAO).				
DE	1) (VAP-1) (HRAO).				
GN	AOC3 OR VAP1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Placenta;				
RX	MEDLINE=97128319; PubMed=8972912;				
RA	Zhang X., McIntire W.S.;				
RT	"Cloning and sequencing of a copper-containing, topaquinine-				
RT	containing monoamine oxidase from human placenta."				
RL	Gene 179:279-286(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.				
RC	TISSUE=Lung;				
RX	MEDLINE=98317014; PubMed=9653080;				
RA	Smith D.J., Salimi M., Bono P., Hellman J., Leu T., Jalilkanen S.;				
RT	"Cloning of vascular adhesion protein 1 reveals a novel				
RT	multifunctional adhesion molecule."				
RL	J. Exp. Med. 188:17-27(1998).				
CC	- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE				
CC	RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO				
CC	PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-				
CC	INDEPENDENT FASHION. HAS A MONOAMINE OXIDASE ACTIVITY.				
CC	- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +				
CC	H(2)O(2).				
CC	- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.				
CC	- SUBUNIT: HOMODIMER.				
CC	- SUBCELLULAR LOCATION: Type II membrane protein.				
CC	- TISSUE SPECIFICITY: MOST STRONGLY EXPRESSED ON THE HIGH				
CC	ENDOTHELIAL VENULES OF PERIPHERAL LYMPH NODES AND ON HEPATIC				
CC	ENDOTHELIA.				
CC	- PTM: Topaquinine (TPQ) is generated by copper-dependent				
CC	autooxidation of a specific tyrosyl residue (By similarity).				
CC	- PTM: N- AND O-GLYCOSYLATED.				
CC	- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL, U39447; AAC50919.1; -				

34 96.5 2.4 1137 1 RIRL_HSV11
35 96 2.3 421 1 SYS_THETH
36 95.5 2.3 553 1 CB62_ARATH
37 95.5 2.3 981 1 EPA3_BRARE
38 95.5 2.3 985 1 AGLU_BOVEY
39 95.5 2.3 2175 1 POLG_BOVEY
40 95 2.3 2258 1 FAS_PIG
41 94.5 2.3 901 1 VEF_GVPU
42 94.5 2.3 1522 1 PSTL_SCHPO
43 94.5 2.3 2555 1 PPS3_BACST
44 94 2.3 492 1 SYTM_YEAST
45 93.5 2.3 878 1 FIMD_ECOLI

P08543 herpes simp
P34945 thernus the
O23066 arabidopsi
O13146 brachydans
P56526 aspergillus
P12915 bovine ente
O941P1 sus scrofa
P41723 pseudolatia
O09750 schizosach
P39847 bacillus su
P48527 saccharomyc
P30130 escherichia


```

DR EMBL: AF067406; AAC25170.1; -.
DR Genew: HGNC:550; AOC3.
DR MIM: 603735; -.
DR InterPro: IPR000269; Cunn_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR Oxioreductase; Copper; TPO; Glycoprotein; Transmembrane;
KM Signal-anchor; Cell adhesion; Polymorphism; Metal-binding.
FT DOMAIN 1 5
FT TRANSMEM 6 26
FT SIGNAL-ANCHOR (TYPE-TI MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 27 763
FT MOD_RES 471 471
FT METAL 520 520
FT METAL 522 522
FT METAL 684 684
FT BINDING 673 673
FT CARBOHYD 137 137
FT CARBOHYD 232 232
FT CARBOHYD 294 294
FT CARBOHYD 592 592
FT CARBOHYD 618 618
FT CARBOHYD 666 666
FT VARIANT 317 317
FT y -> h (IN DBSNP:438287).
FT /FTID-VAR_012064.
SQ SEQUENCE 763 AA; 84621 MW; 58AD55605EC9D228 CRC64;
Query Match 100.0%; Score 4095; DB 1; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.3e-311;
Matches 763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNOKTIVLLILAVITIFALVLCVLLVGRGSGEPSPQLPHCPSPSPSAQPMTHPGQSOLF 60
DB 1 MNOKTIVLLILAVITIFALVLCVLLVGRGSGEPSPQLPHCPSPSPSAQPMTHPGQSOLF 60
QY 61 ADISREELTAVMRELTORGLVDAQAARPSNOCVSVLSOLPRKAAALAHLDRCGSP 120
DB 61 ADISREELTAVMRELTORGLVDAQAARPSNOCVSVLSOLPRKAAALAHLDRCGSP 120
QY 121 AREALAVFEGROPPOPNVSELVGPLPHPSYMRDVYERHGGPLPYRRRVLFOEYDID 180
DB 121 AREALAVFEGROPPOPNVSELVGPLPHPSYMRDVYERHGGPLPYRRRVLFOEYDID 180
QY 181 QMIFNELPQASGLLHHCFFYKRGRLVTMTAPRGLSGDRATWGLYNNISGAGFLL 240
DB 181 QMIFNELPQASGLLHHCFFYKRGRLVTMTAPRGLSGDRATWGLYNNISGAGFLL 240
QY 241 HHVGLLELVNHHKALDPRKMTIQVYOGRTYDSLAOLEAFEGALVNVLLIPDNGTGSW 300
DB 241 HHVGLLELVNHHKALDPRKMTIQVYOGRTYDSLAOLEAFEGALVNVLLIPDNGTGSW 300
QY 301 SLKSPVPFGAPPLQFYPOGPRFSVGSRAVSLMTFSFGLAFSGGRIDVRRGGRV 360
DB 301 SLKSPVPFGAPPLQFYPOGPRFSVGSRAVSLMTFSFGLAFSGGRIDVRRGGRV 360
QY 361 YEISLOEALAIYGSNSPAAMTTRYVDGFGMGKYYTTPRLRGVDCPYLATYVDHFLLESQ 420
DB 361 YEISLOEALAIYGSNSPAAMTTRYVDGFGMGKYYTTPRLRGVDCPYLATYVDHFLLESQ 420
QY 421 APTTIDACVFEQONGRLPRRHSLLSYFEGGLAEVTVLVRSMSLTLLNYDVMDFV 480
DB 421 APTTIDACVFEQONGRLPRRHSLLSYFEGGLAEVTVLVRSMSLTLLNYDVMDFV 480
QY 481 PSQAIRFATGYISSAFEGATGKYGNVSHETLGTVTHSAHFVVDVAVGLEMMV 540
DB 481 PSQAIRFATGYISSAFEGATGKYGNVSHETLGTVTHSAHFVVDVAVGLEMMV 540
QY 541 AEDVVEPAVVPSPHQLOLQVTRKLEMEQAAFLVGSATPRYLVLASNSNKGHP 600
DB 541 AEDVVEPAVVPSPHQLOLQVTRKLEMEQAAFLVGSATPRYLVLASNSNKGHP 600

```

```

DB 541 AEDVVEPAVVPSPHQLOLQVTRKLEMEQAAFLVGSATPRYLVLASNSNKGHP 600
QY 601 RGYRIQMLSFAGEPLLPONSSMARGFSWERYQLAVTORKEEPPSSSVFNONDPAPTVD 660
DB 601 RGYRIQMLSFAGEPLLPONSSMARGFSWERYQLAVTORKEEPPSSSVFNONDPAPTVD 660
QY 661 SDFINNETIAGKDLVAVYTAGFLHIPHAEDIPNTVTVGVNGVGFELRPYNEFDEDPSTY 720
DB 661 SDFINNETIAGKDLVAVYTAGFLHIPHAEDIPNTVTVGVNGVGFELRPYNEFDEDPSTY 720
QY 721 DSIFYRGDDQAGCEVNPPLCLQAAAACADLPFAFISGSGSHN 763
DB 721 DSIFYRGDDQAGCEVNPPLCLQAAAACADLPFAFISGSGSHN 763

RESULT 2
AOC3_MOUSE STANDARD; PRT; 765 AA.
ID AOC3_MOUSE
AC 070423;
AC 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-
DE 1) (YAP-1)
GN AOC3 OR VAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c, and 129/SvJ;
RX MEDLINE=98414290; PubMed=9743358;
RA Bono P., Salmi M., Smith D.J., Leppanen I., Horelli-Kuitunen N.,
RA Palotie A., Jalkanen S.;
RT Isolation, structural characterization, and chromosomal mapping of
RT the mouse vascular adhesion protein-1 gene and promoter.;
J. Immunol. 161:2853-2860(1998).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE
CC RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO
CC PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
CC INDEPENDENT FASHION. HAS A MONOAMINE OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- CORFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (By similarity).
CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB EMBL: AF054831; AAC23747.1; -.
DB EMBL: AF078705; AAC35839.1; -.
DB MGD: MG1:1306797; AOC3.
DR InterPro: IPR000269; Cunn_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
KM Oxioreductase; Copper; TPO; Glycoprotein; Transmembrane;
KM Signal-anchor; Cell adhesion; Metal-binding.
FT DOMAIN 1 6
FT CYTOPLASMIC (POTENTIAL).

```

FT TRANSMEM 7 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL)
 FT DOMAIN 28 765 EXTRACELLULAR (POTENTIAL)
 FT MOD_RES 471 471 TOPADONINE (BY SIMILARITY)
 FT 520 520 COPPER (POTENTIAL)
 FT METAL 522 522 COPPER (POTENTIAL)
 FT 684 684 COPPER (POTENTIAL)
 FT BINDING 673 673 AMILORIDE (BY SIMILARITY)
 FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 659 659 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL)
 SQ SEQUENCE 765 AA; 84533 MW; 74895673DBB44D CRC64;

Query Match 84.8%; Score 3473; DB 1; Length 765;
 Best Local Similarity 83.2%; Pred. No. 4.5e-263;

Matches 634; Conservative 61; Mismatches 67; Indels 0; Gaps 0;

QY 1 MNOKTLLVLLAVITTFALVCLVLRGSGDGEPSQLPHCPSPVSPASQWTHPGSQFL 60
 DB 1 MTKKTLVLLAVITTFALVCLVLRGSGDGEPSQLPHCPSPVSPASQWTHPGSQFL 60
 QY 61 ADLSREELTAVMRFLTORLQGLVDAQAARSPNCVPSVELQIPKRAALAHIDRSGPP 120
 DB 61 ADLSREELTAVMRFLTORLQGLVDAQAARSPNCVPSVELQIPKRAALAHIDRSGPP 120
 QY 121 AREALAVFPPROPPOPNVSELVGLPPLPSPMRDVERHGRGLPYHRRPRLFOEYLDID 180
 DB 121 VREALAVFPPROPPOPNVSELVGLPPLPSPMRDVERHGRGLPYHRRPRLFOEYLDID 180
 QY 181 QMIFNRLPQASGLLHCCCKYKRGKMLVTMTAPRGLOGSRATMFGLYNINSGAFEL 240
 DB 181 EMIFHRELPAQSLHCCCKYKRGKMLVTMTAPRGLOGSRATMFGLYNINSGAFEL 240
 QY 241 HHGELLVNHRKALDPARMTIOKVEYQGRYDLSLAOLEAGLVNVLVPIPNGTGSM 300
 DB 241 HHGELLVNHRKALDPARMTIOKVEYQGRYDLSLAOLEAGLVNVLVPIPNGTGSM 300
 QY 301 SLKSPVPPGAPPLQGYPGGPRFVSQSRVASSLMTFSGLGAFSGPRIFDVRQGERLV 360
 DB 301 SLKSPVPPGAPPLQGYPGGPRFVSQSRVASSLMTFSGLGAFSGPRIFDVRQGERLV 360
 QY 361 YEISLQALAIYGNNSAATRTRYVDGFGMKYTPPLTGVCOPRYLATVDMHFLLESQ 420
 DB 361 YEISLQALAIYGNNSAATRTRYVDGFGMKYTPPLTGVCOPRYLATVDMHFLLESQ 420
 QY 421 APRTIRDAFCVFQNOGLPLRRHSDLYSHYFGLAETVLVVRMSSTLLNVDYVMDTVFH 480
 DB 421 APRTIRDAFCVFQNOGLPLRRHSDLYSHYFGLAETVLVVRMSSTLLNVDYVMDTVFH 480
 QY 481 PSQAIREFYATGYISSAFLEGATGKYGNVSEHTIGTVHTSHAHFKVDDLVAGLENVW 540
 DB 481 PSQAIREFYATGYISSAFLEGATGKYGNVSEHTIGTVHTSHAHFKVDDLVAGLENVW 540
 QY 541 AEMVAVPMAVPSPEHQLOLQVTRKLLMEQOAAFLVSATPRYLYLASNSNKKGHR 600
 DB 541 AEMVAVPMAVPSPEHQLOLQVTRKLLMEQOAAFLVSATPRYLYLASNSNKKGHR 600
 QY 601 RGYRIOMLSPAGEPLPONSMMARGFSEWERYQLAVTORKEEPPSSSVFNONDPAPTYDF 660
 DB 601 RGYRIOMLSPAGEPLPONSMMARGFSEWERYQLAVTORKEEPPSSSVFNONDPAPTYDF 660
 QY 661 SDFINNETIAGKDLVAVNTAGFLHIPHAEDIPMTVTYGVNGVGFELRPYDFEDEDPSFYSA 720
 DB 661 SDFINNETIAGKDLVAVNTAGFLHIPHAEDIPMTVTYGVNGVGFELRPYDFEDEDPSFYSA 720
 QY 721 DSIFYREGDODAGCEVNPCLACLPQAAACADLPFAFSHGFSH 762
 DB 721 DSIFYREGDODAGCEVNPCLACLPQAAACADLPFAFSHGFSH 762

RESULT 3

ACQY_BOVIN
 ID ACQY_BOVIN STANDARD: PRT: 762 AA.
 AC 046406;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Copper amine oxidase, lung isozyme precursor (EC 1.4.3.6) (Amine
 oxidase [copper-containing]) (BOLA0).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_Taxid=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=98151264; PubMed=9492300;
 RA Hogdell E.V.S., Huen G., Borre M., Bundgaard J.R., Larsson L.-I.,
 RA Vuust J.;
 RT "Structure and tissue-specific expression of genes encoding bovine
 RT copper amine oxidases.";
 RL Eur. J. Biochem. 251:320-328(1998).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit (By
 CC similarity).
 CC -1- ENZYME REGULATION: INHIBITED BY AMILORIDE IN A COMPETITIVE MANNER
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Extracellular (By similarity).
 CC -1- TISSUE SPECIFICITY: LUNG, SPLEEN, HEART, KIDNEY.
 CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPADONINE OXIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC
 CC EMBL: Y15774; CA75776.1;
 DR InterPro: IPR000269; CUNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxidn2; 1.
 DR Pfam: PF02728; Cu_amine_oxidn3; 1.
 DR PRINTS: PR00766; CUADOXIDASE.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 DR PROSITE: PS01166; COPPER_AMINE_OXID_3; 1.
 KW Oxidoreductase; Copper; TPQ; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 762
 FT MOD_RES 470 470
 FT METAL 519 519
 FT METAL 521 521
 FT METAL 521 521
 FT METAL 521 521
 FT BINDING 672 672
 FT CARBOHYD 136 136
 FT CARBOHYD 231 231
 FT CARBOHYD 293 293
 FT CARBOHYD 617 617
 FT CARBOHYD 665 665
 SQ SEQUENCE 762 AA; 84883 MW; BB43D04776744AF2 CRC64;

Query Match 81.7%; Score 3346; DB 1; Length 762;

Best Local Similarity 81.9%; Pred. No. 3.6e-253;
 Matches 620; Conservative 46; Mismatches 81; Indels 10; Gaps 2;

QY 14 VITFALVCLVLR--GGDGE-----PSQLPHCPSPVSPASQWTHPGSQFLFADL 63
 DB 3 IFIFLSLWTLVVMGREGGVGSEGVKQCHPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 62

```

QY 64 SRELTAVMFLTORLPGVLVDAAQARPSPDNCVSEVLOLPPKAAALAHLDKRSPPARE 123
DB 63 SRELTAVMFLTORLPGVLVDAAQARPSPDNCVSEVLOLPPKAAALAHLDKRSPPARE 122
QY 124 ALAIVFGRGPOPVNVSELVVGPLPHPSYMDVYVERHGGLPYRRRPVLFQEYLDIDOMI 183
DB 123 ALAIVFGRGPOPVNVSELVVGPLPHPSYMDVYVERHGGLPYRRRPVLFQEYLDIDOMI 182
QY 184 FNNELPOASGLLHHCCEYKRRGRNLVTMTAPRGLSGDRATWFGLYYNSGAGFELAHV 243
DB 183 FNNELPOASGLLHHCCEYKRRGRNLVTMTAPRGLSGDRATWFGLYYNSGAGFELAHV 242
QY 244 GLELVNHRKALDPAKWTIOKVFYGRYDSLAOLEAEFGLVNVVLIIPNGTGSMSLK 303
DB 243 GLELVNHRKALDPAKWTIOKVFYGRYDSLAOLEAEFGLVNVVLIIPNGTGSMSLK 302
QY 304 SPVPGPAPPLQFYPOGPRFSVQSGSVASSLMTFSEGLGAFSGPRIPDFVQERLYEYI 363
DB 303 SPVPGPAPPLQFYPOGPRFSVQSGSVASSLMTFSEGLGAFSGPRIPDFVQERLYEYI 362
QY 364 SLEALAIYGSNPPAAMTTRYVVDGFGMGKYYTPRLTGVCYCPYLATYVDHMFLESOAPK 423
DB 363 SLEALAIYGSNPPAAMTTRYVVDGFGMGKYYTPRLTGVCYCPYLATYVDHMFLESOAPK 422
QY 424 TINDAFCEPQONOGPLRLRHSDLSHYFGSLAETVLRVSMSTLTXDYVMDTVFHPSG 483
DB 423 TINDAFCEPQONOGPLRLRHSDLSHYFGSLAETVLRVSMSTLTXDYVMDTVFHPSG 482
QY 484 AIEIRYATGYISSAFLEGATKYGNVSEHTLGTVHTSHAFKVDLDVAGLEMMVAED 543
DB 483 AIEIRYATGYISSAFLEGATKYGNVSEHTLGTVHTSHAFKVDLDVAGLEMMVAED 542
QY 544 MIVVPVAPVPSPEHQLOLQVTRKLEMEQOAFVLSAIPRYLYXLSNSNKKHGRGY 603
DB 543 MIVVPVAPVPSPEHQLOLQVTRKLEMEQOAFVLSAIPRYLYXLSNSNKKHGRGY 602
QY 604 RIOMLSFAGPPLPONSSMANGFSEWERYQLAVTORKEEPSSSSVFNONDPAFVDPDSF 663
DB 603 RIOMLSFAGPPLPONSSMANGFSEWERYQLAVTORKEEPSSSSVFNONDPAFVDPDSF 662
QY 664 INNETTAKGLVAVMTAGFLHIHAEIDIPNTVGVNGVGFELRPYNFEDDPSFYSDSI 723
DB 663 INNETTAKGLVAVMTAGFLHIHAEIDIPNTVGVNGVGFELRPYNFEDDPSFYSDSI 722
QY 724 YFRGDODAGACEVNHACLPOAAACAPDLPAFSGGF 760
DB 723 YFRGDODAGACEVNHACLPOAAACAPDLPAFSGGF 759

```

RESULT 4 ACOX_BOVIN STANDARD: PRT: 762 AA.

```

AC 029457;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase, liver isozyme precursor (EC 1.4.3.6) (Amine
OS oxidase [copper-containing]) (Serum amine oxidase) (SAO).
OC Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RA MEDLINE=94193686; PubMed=8144587;
RA Nu D., Medzhiridszky K.F., Adams G.W., Mayer P., Hines W.M.,
RA Burlingame A.L., Smith A.J., Cal D., Klman J.P.,
RT "Primary structures for a mammalian cellular and serum copper amine
RT oxidase."
RL J. Biol. Chem. 269:9926-9932(1994).

```

```

RN (2)
RP PARTIAL SEQUENCE.
RX MEDLINE=93090748; PubMed=1457410;
RA Jones S.M., Palcic M.M., Scaman C.H., Smith A.J., Brown D.E.,
RA Dooley D.M., Mure M., Klman J.P.;
RT "Identification of topaquinine and its consensus sequence in copper
RT amine oxidases."
RL Biochemistry 31:12147-12154(1992).
CC 1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC 1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC 1- ENZYME REGULATION: INHIBITED BY AMILORIDE IN A COMPETITIVE MANNER.
CC 1- SUBCELLULAR LOCATION: Extracellular.
CC 1- TISSUE SPECIFICITY: Liver.
CC 1- PTM: Topaquinine (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (by similarity).
CC 1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: S69583; AAB30397.1; -
CC EMBL: L27218; AAA30525.1; -
CC DR InterPro: IPR000269; CuNH_oxidase.
CC DR Pfam: PF01179; Cu_amine_oxid_1.
CC DR Pfam: PF02727; Cu_amine_oxid_2; 1.
CC DR Pfam: PF02728; Cu_amine_oxid_3; 1.
CC DR PRINTS: PRO0766; CUDROXIDASE.
CC DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
CC DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
CC KW Oxidoreductase; Copper; TPQ; Glycoprotein; Signal; Metal-binding.
CC
CC FT CHAIN 1 762 COPPER AMINE OXIDASE, LIVER ISOZYME.
CC FT MOD_RES 470 470 TOPAQUINONE (BY SIMILARITY).
CC FT METAL 519 519 COPPER (POTENTIAL).
CC FT METAL 521 521 COPPER (POTENTIAL).
CC FT METAL 683 683 COPPER (POTENTIAL).
CC FT BINDING 672 672 AMILORIDE (BY SIMILARITY).
CC FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 762 AA; 84756 MW; AA959771360295FE CRC64;

```

Query Match 81.4%; Score 3334; DB 1; Length 762;
Best Local Similarity 80.6%; Pred. No. 3; le-252;
Matches 610; Conservative 61; Mismatches 76; Indels 10; Gaps 2;

```

QY 14 VITIFALVCLVGR--GGDGE-----PSQLPHCPSPVSPAPQWTHPGQSOLFADL 63
DB 3 IFIFLSMTLLVNGREGGVGKQCHPSLPKCPSPSPSDQPTWTHPGQSOLFADL 62
QY 64 SRELTAVMFLTORLPGVLVDAAQARPSPDNCVSEVLOLPPKAAALAHLDKRSPPARE 123
DB 63 SRELTAVMFLTORLPGVLVDAAQARPSPDNCVSEVLOLPPKAAALAHLDKRSPPARE 122
QY 124 ALAIVFGRGPOPVNVSELVVGPLPHPSYMDVYVERHGGLPYRRRPVLFQEYLDIDOMI 183
DB 123 ALAIVFGRGPOPVNVSELVVGPLPHPSYMDVYVERHGGLPYRRRPVLFQEYLDIDOMI 182
QY 184 FNNELPOASGLLHHCCEYKRRGRNLVTMTAPRGLSGDRATWFGLYYNSGAGFELAHV 243
DB 183 FNNELPOASGLLHHCCEYKRRGRNLVTMTAPRGLSGDRATWFGLYYNSGAGFELAHV 242
QY 244 GLELVNHRKALDPAKWTIOKVFYGRYDSLAOLEAEFGLVNVVLIIPNGTGSMSLK 303
DB 243 GLELVNHRKALDPAKWTIOKVFYGRYDSLAOLEAEFGLVNVVLIIPNGTGSMSLK 302
QY 304 SPVPGPAPPLQFYPOGPRFSVQSGSVASSLMTFSEGLGAFSGPRIPDFVQERLYEYI 363

```

Db 303 SOVPEPTPLPQHPGPFPSVGNVASSLMTFSGLGFSGPRFVDVAFQGERLAYET 362
 Qy 364 SLOEALAIYGNSSPAAMTTRVYDVGFGMGKRTPLTRGVDCPYLATYVDMHFLLESQAPK 423
 Db 363 SLOEAGAVYGNTPAAMLTFRYDMSGFMGFATPLLRGVDCPYLATYVDMHFLVESQTPK 422
 Qy 424 TIDAFCEVQONQGLPLRRHSDLYSHYFGGLAEVLYVRSKSTLNTLYDMTVFHPSG 483
 Db 423 TLHDACVEQONKGLPLRRHSDLYSHYFGGVAQTVLVERSVSTMLNYDYVMQVFPNG 482
 Qy 484 AIRIRYATGYISSAFLEFGATGYGNVSEHTGTHTSHAHKRVOLDVAGLNMWMAED 543
 Db 483 AIEVKLHATGYISSAFLEFGAARRKGNVGHETLGPVHTSHAHKRVOLDVAGLNMWMAED 542
 Qy 544 MVEFPAVWSPERHQLQRLQVTRKLLMEMBOAFLVGSATPRYLYLASNSHNRKMGHGRGY 603
 Db 543 MAFVPAIAPWSPERHQLQRLQVTRKLLMEMBOAFLVGSATPRYLYLASNSHNRKMGHGRGY 602
 Qy 604 RIOMLSFAGEPLPQNSSMARGESWERYQLAVTORKEBESSSVFNONDPAATVDSDF 663
 Db 603 RIQTVSEFAGPMPQNSPMEAFSGRQLAITQRKETEPSSSVFNONDPAATVDSDF 662
 Qy 664 INNETIAGKDLVAMTAGFLHPIHAEIDPMTVTVGNVGFELRPYNFDEDPSPYSADSI 723
 Db 663 INNETIAGKDLVAMTAGFLHPIHAEIDPMTVTVGNVGFELRPYNFDEDPSPYSADSI 722
 Qy 724 YFRGDDAGACEVNPPLACLPQAACAADLPFASHGGE 760
 Db 723 YFRGDDAGACEVNPPLACLPQAACAADLPFASHGGE 759
 RESULT 5
 AOC2_HUMAN
 ID AOC2_HUMAN STANDARD: PRT: 756 AA.
 AC 075106; 075105; 000120; 09UNY0;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Retina-specific copper amine oxidase precursor (EC 1.4.3.6) (RAO)
 DE (Amine oxidase [copper-containing]).
 GN AOC2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RX MEDLINE=97237047; PubMed=9119395;
 RA Imanura Y., Kubota R., Wang Y., Asakawa S., Kudoh J., Mashima Y.,
 RA Oguchi Y., Shimizu N.;
 RT "Human retina-specific amine oxidase (RAO): cDNA cloning, tissue
 RT expression, and chromosomal mapping.";
 RL Genomics 40:277-283(1997).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=98390194; PubMed=9722934;
 RA Imanura Y., Noda S., Mashima Y., Kudoh J., Oguchi Y., Shimizu N.;
 RT "Human retina-specific amine oxidase: genomic structure of the gene
 RT (AOC2), alternatively spliced variant, and mRNA expression in
 RT retina";
 RL Genomics 51:293-298(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RA Zhang X., McInliffe W.S.;
 RT "Human copper-containing amine oxidases";
 RL Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE A CRITICAL MODULATOR OF SIGNAL TRANSMISSION IN
 CC RETINA, POSSIBLY BY DEGRADING THE BIOGENIC AMINES DOPAMINE,
 CC HISTAMINE, AND PUTRESCINE.
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +

CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaguinone per subunit.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: RETINAL-SPECIFIC.
 CC -1- PTM: Topaguinone (TPQ) is generated by copper-dependent
 CC autoxidation of a specific tyrosyl residue (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 DR EMBL: AB012943; BAA32590.1;
 DR EMBL: AB012942; BAA32590.1; JOINED.
 DR EMBL: AB012943; BAA32589.1;
 DR EMBL: AB012942; BAA32589.1; JOINED.
 DR EMBL: D88213; BAA19001.1;
 DR EMBL: AF081363; AAD39345.1;
 DR Genew: HGNC:549; AOC2.
 DR MIM: 602268;
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxidn2; 1.
 DR Pfam: PF02728; Cu_amine_oxidn3; 1.
 DR PRINTS: PR00766; CUDAOXIDASE.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 DR Oxidoreductase; Copper; TPQ; Alternative splicing; Glycoprotein;
 KW Signal; Metal-binding.
 KM
 FT SIGNAL 1 32
 FT CHAIN 33 756
 FT MOD_RES 465 465
 FT METAL 516 516
 FT METAL 518 518
 FT METAL 680 680
 FT BINDING 669 669
 FT CARBOHYD 133 133
 FT CARBOHYD 198 198
 FT CARBOHYD 226 226
 FT CARBOHYD 388 388
 FT CARBOHYD 662 662
 FT VARSPLIC 599 625
 FT CONFLICT 181 181
 FT CONFLICT 215 218
 FT CONFLICT 221 222
 FT CONFLICT 610 610
 SQ SEQUENCE 756 AA; 83793 MW; 4E3B7317E6DCA66 CRC64;
 Query Match 63.8%; Score 2612; DB 1; Length 756;
 Best Local Similarity 65.0%; Pred. No. 6; 6e-196;
 Matches 495; Conservative 92; Mismatches 167; Indels 8; Gaps 3;
 Qy 1 MNOKTILLVLLAVITLALVCLVGVGGDGGESQLPHCPSPVSAQPFTHGQSOLF 60
 Db 1 MHLKIYAFALSLITLIFALAYVLLTSPGSS----SQPHCPVSHRAQWPHRGQSOLF 56
 Qy 61 ADSRELLTAVMFLTRQLRGLVDAQAQARSDNCVSVELQLPKKAALAHDRGSP 120
 Db 57 ADSRELLTAVMFLTRQLRGLVDAQAQARSDNCVSVELQLPKKAALAHDRGSP 116
 Qy 121 AREALATVFGROPQPVNSLVVGLPSPYMRDVTERRGGPLPYRRPVLFQETDID 180
 Db 117 AREALATVFGGQPOPVNSLVVGLPSPYMRDVTERRGGPLPYRRPVLAELFQM 176
 Qy 181 QMIFNRELPOASGLLHHCCEYKRRGNLVTMTAPRGLOGSDRATWGLYINISGAGFL 240
 Db 177 RHLDKVELPKRAPIFLST--FNYNGSTLAAVHAHPRLRSRERTYMAALYHNISGVCLF 234

Query Match	34.2%	Score 1400.5;	DB 1;	Length 746;
Best Local Similarity	39.1%	Pred. No. 2.2e-101;		
Matches 300; Conservative 131; Mismatches 280;			Indels 57; Gaps 14	

D	b	51	REELGLDQPSKEPLTANKSVLEIMLPKKKHVLKFLDEGRKQGNPREARAVIFEGQADY	110
Q	y	138	VSELYLVGRPLPHSPYMDVYVERHGGLPLPHRRPVYLEQLELDDIMQTFNELQOASGLLH	197
D	b	111	VTEFAPVGLPRPYIRLRS-PRGHNHLSWSRPISTAEY---DLLYH-TLKRATPMLQ	164
Q	y	198	-----CCFYKRGRLVMTTAPRGLOSQDATYFGLYTNISGAGFFLHVGLLELVN	250
D	b	165	FLDDTTGFSFLGDDDLTTTVDAPRGVASGQSRSMFYIQRYVE-GYLLHPTGLEILLD	222
Q	y	251	HKALDPARKWTIQVYQGRYYSLQALEQAFEGALNVVLLDP--NGTGSWSLKSPPV	306
D	b	223	HGSTVDQDMVEEDLWYNGKFFYNNPEELARKYUAAVEQDVTJVLDEPLRPNGTE---KPLP	277
Q	y	307	-----PPGAPRLDLYPOGCRFESQGRVASSLMTFSPFGIAGFSGRIRVDVQ	355
D	b	278	SSYKPRGEFTPPVNAAGPHVQPSGRPKLEGNTVLYGKNSFSYLRSSSGQITRNVLFG	337
Q	y	356	GERLVEIISLQELAIYGSNPAAMTTRVYDGGFGMKYTPLTRGDCPLATYVDHMF	415
D	b	338	GERVAYEVSYQEAVALYGHTRPAGMOTKYIDVGMGLSVTHELAPGIDCPEATFLDAFH	397
Q	y	416	LLSQAKTIRDAFCVFEQNGQLPLRRHSDLS---HFGGIAEYLVLYVRMSITLND	472
D	b	398	YVSDGQVHYPHALCFEEMPTGVLPRHNFNSNKGGFNFYAGLKGVLALRTSTVYND	457
Q	y	473	YVMDTVPHSGAIEIRFYARGYISSAFLEFGATKYGUNQSEHTLGTVHSHAFKVDLY	532
D	b	458	YIMDFIYNSGVMEAKHAKGYHAFYFPEGRJGHTRLQTHLGNHRLVHYRDMQY	517
Q	y	533	AGLENWYAEADWYFVPRAVWSPFHOLQRLQVTRKLLMEBOAFLVGSATRTLYLASN	592
D	b	518	AGTNSFQTLTKLENTJNWSMSSHSLYQPTLQOTYSOEHOAAFFQGTLEKYLFFSSP	577
Q	y	593	HSNKGHPRGYRIOMLSFACEPLPONSMSARGFSEMERYOLATVQRKEEPPSSSVENQD	652
D	b	578	QKNQMGHRRSRYRLQIHSMAGQVLPQMGQERATYMARPLATYKRYESERYSSSLXND	637
Q	y	653	PWATVDFSDFI-NNETIAGKDLVAVYTAGFLHIIPAEDIPNTVTVGNGVGFLLRPYEF	711
D	b	638	PMPBPVYFEELFRNNENIEDEDLVAVYTGFLHIIHSEEDVPNTATGNSVGLLRPFNF	697
Q	y	712	DEDSFYSADSIYFRGDQDAGACEVNPALQDPAACAPDLPAPFSGG	759
D	b	698	PEDSLASRDTVIV-WPDQGLNRVQ--RWIPDRCLVVS-PPFSYNG	741
RESULT 8				
ABP_HUMAN				
ID	ABP_HUMAN	STANDARD:	PRT:	751 AA.
AC	P19801: Q16683: Q16684:			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Amloride-sensitive amine oxidase [copper-containing] precursor			
DE	(EC 1.4.3.6) (diamine oxidase) (DAO) (Amloride-binding protein)			
DE	(ABP) (Histaminase).			
GN	ABP1 OR DAO1 OR AOCL1.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Choriata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94237856; PubMed=8182053;			
RA	Chassande O., Renard S., Barbry P., Iazdunski M.;			
RT	"The human gene for diamine oxidase, an amiloride binding protein.			
RT	Molecular cloning, sequencing, and characterization of the promoter."			
RL	J. Biol. Chem. 269:14484-14489(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=96113540; PubMed=8595053;			

RA Zhang X., Kim J., McIntire W.S.;
 RT "cDNA sequences of variant forms of human placenta diamine oxidase."
 RL Biochem. Genet. 33:261-268(1995).
 RN [3]
 RP PRELIMINARY SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-KIDNEY;
 RX MEDLINE-91017502; PubMed-2217167;
 RA Barbry P., Champe M., Chassande O., Munemitsu S., Champigny G.,
 RA Lingueglia E., Maes P., Frelin C., Tartar A., Ullrich A.,
 RA Lazdunski M.,
 RT "Human kidney amiloride-binding protein: cDNA structure and functional
 expression".
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7347-7351(1990).
 RN [4]
 RP SEQUENCE OF 20-39, AND CHARACTERIZATION.
 RC TISSUE-Placenta.
 RX MEDLINE-94193685; PubMed-8144586;
 RA Novotny W.F., Chassande O., Baker M., Lazdunski M., Barbry P.;
 RT "Diamine oxidase is the amiloride-binding protein and is inhibited by
 amiloride analogues".
 RL J. Biol. Chem. 269:9921-9925(1994).
 CC -1- FUNCTION: CATALYZES THE DEGRADATION OF COMPOUNDS SUCH AS
 PUTRESCINE, HISTAMINE, SPERMINE, AND SPERIDINE, SUBSTANCES
 INVOLVED IN ALLERGIC AND IMMUNE RESPONSES, CELL PROLIFERATION,
 TISSUE DIFFERENTIATION, TUMOR FORMATION, AND POSSIBLY APOPTOSIS.
 CC PLACENTAL DAO IS THOUGHT TO PLAY A ROLE IN THE REGULATION OF THE
 CC FEMALE REPRODUCTIVE FUNCTION.
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
 H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: PLACENTA AND KIDNEY.
 CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent
 autoxidation of a specific tyrosyl residue.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X78212; CA55046.1; -;
 DR EMBL: U11863; AAC50270.1; -;
 DR EMBL: U11863; AAB60381.1; -;
 DR EMBL: M55602; AAA58358.1; ALT_SEQ.
 DR PIR: A38276; A38276.
 DR PIR: S42495; S42495.
 DR GeneW: HGNC:80; ABP1.
 DR MIM: 104610; -;
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid. 1.
 DR Pfam: PF02727; Cu_amine_oxidn3; 1.
 DR Pfam: PF02728; Cu_amine_oxidn3; 1.
 DR PRINTS: PR00766; CUDAOXIDASE.
 DR PROSITE: PS01164; COPPER AMINE OXID. 1.
 DR PROSITE: PS01165; COPPER AMINE OXID. 2; 1.
 DR Signal: glycoprotein; Oxidoreductase; Copper; Heparin-binding; TPQ;
 KW Alternative splicing; Polymorphism; Metal-binding.
 FT SIGNAL 1
 FT CHAIN 20 751
 FT METAL 391 391
 FT MOD_RES 461 461
 FT METAL 510 510
 FT METAL 512 512
 FT METAL 675 675
 FT BINDING 568 575
 FT BINDING 664 664
 FT CARBOHYD 110 110
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 538 538 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 619 619 R -> RTGGQPALSOARSPVGR (IN ISOFORM
 FT VARIANT 645 645 DAO2).
 FT VARIANT 645 645 H -> D.
 FT CONFLICT 28 28 /FTID=VAR. 007542.
 FT CONFLICT 332 332 R -> A (IN REF. 2).
 FT CONFLICT 332 332 S -> F (IN REF. 2).
 SO SEQUENCE 751 AA; 85363 MW; 1B8B695C44232DA CRC64;
 Query Match 34.0%; Score 1394; DB 1; Length 751;
 Best Local Similarity 39.8%; Pred. No. 7.1e-101;
 Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;
 50 PWTGQSOLFADLSREBLVNMFLTORGLGVDAQAARSPNCVPSVELDPPRAAA 109
 23 PGLTPRACGFSDLSNOLKAVHSFLWSKELRLQPSSTTMANTVFLEMLPKKYHV 82
 110 LAHLDRGPPRAREALVIFGROPOPNSVSLVGPPLPHEPMYMDVYVERHGCLPYNHR 169
 83 LRFIDKGRHFRARAVIFGDEGDEHNVTEFAVGPLPGCYMALSPRGYSSMASR 141
 170 PVLFOEYLDIDOMTFNRELPOASGLNH-----CCFYKRGHNLVMTTAPRGLOSGD 222
 142 PISTAEY-----ALLYHTLOEATKPLHQFNTGTFSPQDCHRCIAFTDVAARGVASQ 196
 223 RATPFGLYNLSGAGFLNHNGLVNLNKAALDPAKRTQKVFQGRYDSLAOLEQFE 282
 197 RRSMLIIORYVE-GYFLHPGLLELDVHGSTDGHNVAVEDVWYNGKFGSPLEAKRYA 254
 283 AGLVNVLIPNGTGGSSLSKSPVP-----GPAP-----PLQFYQGRFSGV 327
 255 DGEVDVVLLEPLDPLGCKHDSDEPPLFSHKRPGDPSPPLHVGPRVLDVPHGRFRLEG 314
 328 SRVASSLWTFSGFGLAASGPRIFDVROGERLVEISLQELALAYGNSPAAMTTRYVDG 387
 315 NAVLYGGSFARLRSSGLQVLNVHGEGRVIAEVSQEAVALYGGHTPAGMGTXYLD 374
 388 GFGGKGYTPPLTRGVDCYLATYVDHMLLESQAKPIRDAFCVFEQNOGLPLRRHNSDL 447
 375 GWSGLSVTHLAPIDCEPDTFLDTFHYDADPVHPRALCFEEMPTGVPLRRHNSN 434
 448 YS---HYFGGLAEVNLVVRMSSTLNDYVVDVTFHSGAIEIFYATGYSSAFEGAT 504
 435 FKGGFNFAAGLKGVLVLRITSTYVNDYINDYIFEPNGVNAEMHNTGYNAHFTYREG 494
 505 GRYGNVSEHTLGVTHNSAHFKYDLDVAGLENNVMAEDVFEVMAVWPSPHOLQRLQV 564
 495 LRHGTRLTHLIGHIHTLVHYRVDLVAGTKNSFOTLQMKLENTITNWSRHHRYVPTL 554
 565 TRKLEMEQOAFVLGSAATPRYLVLASHSNKGWHPRGYRIOMLSFAGEPLPQUSNARG 624
 555 EOTYSWERQAFRRKRLPKRYLLFTSPQENPMWCHKRSYRLQIHSMDQVLPRPMQEBQA 614
 625 FSWERYQLAVTQREKEEPESSSSVNONDPAPATVDFSDFI--NNETIAGKDLVAVATAGFL 683
 615 ITMARVPLAVTKYESSELSSSIYHONDPMPRVVFEGLNNNNENIEDLVAVATVGL 674
 684 HIPHAEDIPNTVGVNGVGFPLRPYNFEDEPSTYSADSIYFRCDQDAGACEVNPPLACL 743
 675 HIPSEDPNATPGNSGVFLRPENFEDEPDSIASRDTIV-TPRNGPMYVO--RWIP 731
 744 QAAACAPDLPFAFSGHG 759
 732 EDRDCSMP-PPFSTNG 746
 DB 732 EDRDCSMP-PPFSTNG 746
 RESULT 9
 AMO_KLEAE STANDARD; PRI: 755 AA.
 ID AMO_KLEAE
 AC P49250;
 DT 01-FEB-1996 (Rel. 33, Created)


```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase precursor (EC 1.4.3.6) (Monamine oxidase)
DE (Tyramine oxidase).
GN MAOA OR TYNA.
OS Klebsiella aerogenes.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxId=28451;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC STRAIN=M70;
RX MEDLINE=92210491; PubMed=1556068;
RA Sugino H., Sasaki M., Azakami H., Yamashita M., Murooka Y.;
RT "A monamine-regulated Klebsiella aerogenes operon containing the
RL monamine oxidase structural gene (maoa) and the maoc gene.";
J. Bacteriol. 174:2485-2492(1992).
CC -1- FUNCTION: ACTIVE ON TYRAMINE, TRYPTAMINE, BETA-PHENETHYLAMINE AND
CC DOPAMINE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaguinone per subunit.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: BY TYRAMINE AND CATECHOLAMINES.
CC -1- PTM: Topaguinone (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (by similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10208; BAA01060.1; -
DR HSSP: P46883; 10AC.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR OXidoreductase; Copper; TPQ; Periplasmic; Signal; Metal-binding.
FT CHAIN 1 755 COPPER AMINE OXIDASE.
FT ACT SITE 413 413 CATALYTIC BASE (BY SIMILARITY).
FT MOD RES 496 496 TOPAGUINONE (BY SIMILARITY).
FT METAL 554 554 COPPER (POTENTIAL).
FT METAL 556 556 COPPER (POTENTIAL).
FT METAL 719 719 COPPER (POTENTIAL).
SQ SEQUENCE 755 AA; 83576 MW; 7B5552283CD93EFC CRC64;

Query Match 10.1%; Score 413; DB 1; Length 755;
Best Local Similarity 27.5%; Pred. No. 2.5e-24;
Matches 150; Conservative 75; Mismatches 213; Indels 108; Gaps 21;

```

```

OY 442 RHHSLSHYFGG-----LAETLVVRSKSTLLNDYVDVTPHPSGATIRFYATGIS 486
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase precursor (EC 1.4.3.6) (Monamine oxidase)
DE (Tyramine oxidase).
GN MAOA OR TYNA.
OS Klebsiella aerogenes.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxId=28451;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC STRAIN=M70;
RX MEDLINE=92210491; PubMed=1556068;
RA Sugino H., Sasaki M., Azakami H., Yamashita M., Murooka Y.;
RT "A monamine-regulated Klebsiella aerogenes operon containing the
RL monamine oxidase structural gene (maoa) and the maoc gene.";
J. Bacteriol. 174:2485-2492(1992).
CC -1- FUNCTION: ACTIVE ON TYRAMINE, TRYPTAMINE, BETA-PHENETHYLAMINE AND
CC DOPAMINE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaguinone per subunit.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: BY TYRAMINE AND CATECHOLAMINES.
CC -1- PTM: Topaguinone (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (by similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10208; BAA01060.1; -
DR HSSP: P46883; 10AC.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR OXidoreductase; Copper; TPQ; Periplasmic; Signal; Metal-binding.
FT CHAIN 1 755 COPPER AMINE OXIDASE.
FT ACT SITE 413 413 CATALYTIC BASE (BY SIMILARITY).
FT MOD RES 496 496 TOPAGUINONE (BY SIMILARITY).
FT METAL 554 554 COPPER (POTENTIAL).
FT METAL 556 556 COPPER (POTENTIAL).
FT METAL 719 719 COPPER (POTENTIAL).
SQ SEQUENCE 755 AA; 83576 MW; 7B5552283CD93EFC CRC64;

Query Match 10.1%; Score 413; DB 1; Length 755;
Best Local Similarity 27.5%; Pred. No. 2.5e-24;
Matches 150; Conservative 75; Mismatches 213; Indels 108; Gaps 21;

```



```

CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- PM: Topaguinone (TPQ) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue.
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D23670; BAA03490.1; -
DR EMBL; L47571; AAC37012.1; -
DR EMBL; AE000235; AAC34468.1; -
DR EMBL; D90777; BAA14996.1; -
DR EMBL; D90776; BAA14993.1; -
DR EMBL; X97452; CAA66104.1; -
DR EMBL; X97453; CAA66107.1; -
DR PDB; 1OAC; 03-APR-96.
DR PDB; 1SPU; 12-MAR-97.
DR PDB; 1QAF; 23-AUG-99.
DR PDB; 1QAK; 24-AUG-99.
DR PDB; 1QAL; 24-AUG-99.
DR PDB; 1DYU; 29-FEB-00.
DR PDB; 1D6Y; 02-FEB-00.
DR PDB; 1D6Y; 02-FEB-00.
DR PDB; 1D6Z; 02-FEB-00.
DR Ecocore; EG13140; tYNA.
DR InterPro; IPR000269; CuNH_oxidase.
DR Pfam; PF01179; Cu_amine_oxid. 1.
DR Pfam; PF02727; Cu_amine_oxidN3; 1.
DR Pfam; PF02728; Cu_amine_oxidN3; 1.
DR PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE; PS01165; COPPER_AMINE_OXID_2; 1.
KW Oxidoreductase; Metal-binding; Copper; Calcium-binding; TPQ;
KW Periplasmic; Signal; 3D-structure; Complete proteome.
KW SIGNAL
FT CHAIN 1 30
FT CA_BIND 563 563 COPPER AMINE OXIDASE.
FT CA_BIND 564 564 1 (VIA CARBONYL OXYGEN).
FT CA_BIND 565 565 1.
FT CA_BIND 708 708 1.
FT CA_BIND 709 709 1 (VIA CARBONYL OXYGEN).
FT CA_BIND 603 603 2.
FT CA_BIND 700 700 2.
FT CA_BIND 702 702 2.
FT CA_BIND 709 709 2 (VIA CARBONYL OXYGEN).
FT ACT_SITE 413 413 CATALYTIC BASE (PROBABLE).
FT MOD_RES 496 496 TOPAGUINONE.
FT METAL 554 554 COPPER.
FT METAL 556 556 COPPER.
FT METAL 719 719 COPPER.
FT VARIAM 42 42 K -> E (IN STRAIN W).
FT VARIAM 59 59 L -> I (IN STRAIN W).
FT CONFLICT 33 33 G -> E (IN REF. 7).
FT CONFLICT 248 248 K -> E (IN REF. 1).
FT CONFLICT 258 259 GY -> VI (IN REF. 1).
FT CONFLICT 276 276 I -> II (IN REF. 1).
FT CONFLICT 288 288 MISSING (IN REF. 1).
FT CONFLICT 290 290 P -> I (IN REF. 1).
FT CONFLICT 456 456 A -> P (IN REF. 1).
FT CONFLICT 659 659 H -> D (IN REF. 1).
SQ SEQUENCE 757 AA; 84378 MW; 65600BCED35243DB CRC64;

Query Match 9.7%; Score 397.5; DB 1; Length 757;
Best Local Similarity 27.5%; Pred. No. 4e-23;
Matches 142; Conservative 74; Mismatches 211; Indels 89; Gaps 20;

265 FYQSGHYDLSLAQLEAFGLVNVVLIPDNGTGSWS-----LKS 304
259 YEPDK--DGLDQ-----DAPILKVIYSYDVGDNVWADPTENIVAVVYLPKRTVYKFFG 311

```

```

QY 305 PVPKGP-----AP---PLOGY-POGPRFSVOGSRVASSLWTFSGGLAFSGPR 348
DB 312 PVPVPMPTARPFDRDRDAVAVPMQIIEEGKNYITGDMIHRRNDHLSNMSRVGPM 371
QY 349 IFVPRFG-----BRLVYELQELALAYGNSPAAATRRVYDG-FGNGKTTPLTGLVD 403
DB 372 ISVTYINDNCKTKKVMESLGCNTVPGDIDGWFKALDSDGYMGTLSPDIARGKD 431
QY 404 CPYLATYVDHMFLESQAPRTINDAFCEPNOGCLPLRRHSDLSYFPG-----LAET 458
DB 432 AFSNAVLNETIADYGVPMERAIKAVFE-----KAGPEYKQEMQGPVSTERR 483
QY 459 VLVYRSKSTLNTDYWDVYFHPGSAIEIRFATGY-----ISSAFLEGATGK---YGN 509
DB 484 ELVYRMWISTGVNDYIFDMIFHENGITIGIDAGATGEAVKGVAKTMDDETAKDTRGYT 543
QY 510 QVSEHTLGTVHTSAHKVLDVAGLENNYMAEDMVFVPAVVRSPHOLQRLDYTKKL 569
DB 544 LIHNTVGTTHOITVNERLDDVDGENNSLVAMDVVKPMT---AGGPRSTJMVQNOYNI 600
QY 570 EMEQOAFVLSATPRILYLASHNSNMKGHPRGYRIQMLSFAG--EPLPONSSMA----- 622
DB 601 GNBQDAQKDFPGTIR-LISNPKENMGNPVS--QIIRYAGTHVAAQAPAPDEMI 657
QY 623 -RQFSERVOLANTOKREEPPSSSVF-NONDPAPTVDFSPINNETTAGKDLVAVTA 680
DB 658 YHRLSFMDKOLWYTRHGHGRPEGKYPNRSTHDTGLGYSK--DNESLNTDAVVMWT 715
QY 681 GFLLIHPAEDIPNTVTVNGVGFELRYPNFEDPS 716
DB 716 GTTHVARAEEMPTMPT--EMVHTLLKPMNFDEPT 749

RESULT 11
AMOH_ARTGO STANDARD: PRT: 683 AA.
AC 059118:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histamine oxidase (EC 1.4.3.6) (Copper amine oxidase).
OS Archaeobacter globiformis.
OC Bacteria: Actinobacteria: Actinobacteria (class): Actinobacteridae;
OC Actinomycetales; Micrococccineae; Micrococaceae; Arthrobacter.
OX NCBI:taxid=165;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-IFO 12137 / ATCC 8010;
RX MEDLINE-95181469; PubMed-7876243;
RA Choi Y.-H., Matsuzaki R., Fukui T., Shimizu E., Yoritani T., Sato H.,
RA Ozaki Y., Tanizawa K.;
RT Copper/topa quinone-containing histamine oxidase from Arthrobacter
RT globiformis. Molecular cloning and sequencing, overproduction of
RT precursor enzyme, and generation of topa quinone cofactor.";
RL J. Biol. Chem. 270:4712-4720(1995).
CC -1- FUNCTION: OXIDIZE HISTAMINE. OTHER AMINES INCLUDING
CC PHEETHYLAMINE, TYRAMINE, TRYPTAMINE, PUTRESCINE, AND BENZYLAMINE
CC ALSO SERVE AS SUBSTRATE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: BY HISTAMINE.
CC -1- PTM: Topaquinone (TPQ) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue.
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DB EMBL: D38508; BA07517.1; -.
DR HSSP: P46881; 1A44.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid. 1.
DR Pfam: PF02727; Cu_amine_oxid2. 1.
DR Pfam: PF02728; Cu_amine_oxid3. 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID. 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID. 2; FALSE_NEG.
KW Oxidoreductase; Copper; TPQ; Metal-binding.
FT INIT_MET 0
FT MOD_RES 401 401 TOPAQUINONE.
FT METAL 450 450 COPPER (POTENTIAL).
FT METAL 452 452 COPPER (POTENTIAL).
FT METAL 611 611 COPPER (POTENTIAL).
SQ SEQUENCE 683 AA; 74979 MW; 233C46C017201AA0 CRC64;

Query Match 9.2%; Score 377; DB 1; Length 683;
Best Local Similarity 24.1%; Pred. No. 1,4e-21;
Matches 182; Conservative 107; Mismatches 343; Indels 124; Gaps 28;

QY 39 PRCPSVPSAQPMT--HPGSQLFADLSREELTAVMRFLTORLPGLVDAQAQRPDNCV 96
DB 6 PSTPLVQDDPPVPATLVHAAQHPLEQSAEIHERRILAE--AGLVG----- 51
QY 97 FSVELQPPRAALAHNDRSP-----PARALAVFFGRPO-----PVSE 140
DB 52 -----ESTRPAYGLIEPPTTRQGVGTGAARLVRLMADMAOSRLDVRSLATG 102
QY 141 LVVGPLPSPYMRDVTYERHGGPLVHRRVPLFOEYLDIDOMIFNELPOASGLHHCCE 200
DB 103 LVVD-----RRELNEAG-----QLVLEEFQIIEIDIL--SEDPQNNAL----- 142
QY 201 YKHGRNLVTMTAP-----RGLSGDRATWFGLYNISGAGFLHNVGLLELVNKA 253
DB 143 -TAGLTPACQVRAVPLSAGVFEYGENEGKRL-----LRGLGFQDDHP-----ADHPW 188
QY 254 LDPARMTIOKVFQGRYDLSLAQLEQFEGGLVNVVLPD--NCTGGSMSLSKVPPOPAR 312
DB 189 AHFIDGLVAVDVENRRVNL-----IDGSPVPV--PEVNGYITDPAIRGELRTDLP 239
QY 313 PLOGYPOGPRFSVOGSRVASSLWTFSGGLAFSGPRIFDVR--FOGER--LYVEISLOEA 368
DB 240 IEIMOPEGSPFTLEGNHLSMAGMDLRVGFPAARGGLVHQHSHKGRRRVIRASISEM 299
QY 369 LAIYGNSPAAATRRVYDG-FGNGKTTPLTRGVDCPYLATYVDHMFLESQAPKTIRD 427
DB 300 VVEYGPSPYRSMQNYFDSGEYLVRDANSLRGCDLGDITVMSPVVADDFGNPTREN 359
QY 428 ACFVFPONOGCLPLRRHSDLSYFPGSLATVYLVNMSGLLVWDVYWDVYFHPGSAIEI 487
DB 360 GTIHEEDAGILMK--HTDEMAQSDYERRRRLVSEFTTVGVYDGFYUWYLLDGTIER 417
QY 488 RFYATGYSASFLEGATGKGNQVSEHTLGTVHTSAHFQVDDVAGLENNYMAEDVVF 547
DB 418 EAKATGIVFTALPDDVAVASIELAGLCAPIYQHILFSARLDMIDGADRVELDLVRL 477
QY 548 PNAVVPSPHOLQRLDYTKKLEMEQOAFVLSATPRILYLASHNS-NKMGHPRGYRIQ 606
DB 478 PKG-PGNPHG--NAFTQKRTLLARESEAVVDADGAGRGVWHISNPDSLNHLGHPGYTL 534
QY 607 MTSFAGEP---LPONSSMARGFSEMYVOLAATVORKEEPPSSSVFNONDPMAATVDFSD 663
DB 535 P--EENPTLLAMDDSSISRAAFARHHLVTRHAAEEELYAADDVFNQHGAVLP-AYV 590
QY 664 INNETIAGDLVAVMTAGFLIHPHADIPTVTVNGVGFELRYPNFEDPSFYGASDI 723
DB 591 AQRDIDGDLGVVHSHFGLHFRPREDWP--INPVTTGFTLKPHEFFENDPTLVNPSA 648
QY 724 YFSGDDAGACEVNPACLPQAAACAPDLPATSHGG 759

```

Db 649 -----AGHGTG-----SERENADGTAHVSHG 672

RESULT 12

AMOI_ARTS1 STANDARD; PRT; 648 AA.

AC 007121;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE COPPER amine oxidase precursor (EC 1.4.3.6) (MAOXI).

GN MAOI.

OS *Arthrobacter* sp. (strain P1).

OC Bacteria; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Micrococcales; Micrococcaceae; *Arthrobacter*.

OX NCBI_TaxID=47915;

RN [1]

RP MEDLINE=93374858; PubMed=8366046;

RX Zhang X., Fuller J.H., McIntire W.S.;

RT Cloning, sequencing, expression, and regulation of the structural

RT gene for the copper/topa quinone-containing methylamine oxidase from

RT *Arthrobacter* strain P1, a Gram-positive facultative methylotroph.;

RL J. Bacteriol. 175:5617-5627(1993).

CC -1- FUNCTION: THE EXACT FUNCTION OF MAOXI IS NOT KNOWN.

CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

CC H(2)O(2).

CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinoxone per subunit.

CC -1- PTM: Topaquinoxone (TPO) is generated by copper-dependent

CC autooxidation of a specific tyrosyl residue (by similarity).

CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAUQUINOXONE OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC

CC

CC

CC

CC

QY 232 NISGAFELHHVGLLELVNHNKALD-PARMTIQK--VEYQGRYDSLQAEQFAGLVN 288

Db 149 GEDNEGRRL-MRALVVRDEADSPYAHPIENIVY-----DLNAG--KV 191

QY 289 VLIPDNCTGSGWSLSPVPPGAPPIQFY-----PGGPRSVGGSRVASS 333

Db 192 VRLDDQA-----IPPSARGNYLPRKYVEARTDLKPLNITQPEGASFVGNHTWA 244

QY 334 LMTSPFGAGSSGRIDVRQGE-----RLVYEISLQELALYGNSPA-AMTRYVDG 388

Db 245 DMSFRVGTTPREGVLVQLKFKDGVDRPVYINRSLSEMVVYDGTAPVQKKNAFSGE 304

QY 389 FGMRKTTPLRTGVDCPYLATYVDHMFLESOAKTIRDAFCVEQNOGLPLRRHSDLY 448

Db 305 YNIGMANSLTGLDCCGEIYFPGHVSVDNGNWTIENALCMEDDST-----LM 356

QY 449 SH--YFGGLAET---VLVRSMTLLNYDVMVTVPHPSCAIEIRYANGYISA-FLF 501

Db 357 KHDFREGTAETRRSRKLVISFIATVANYEAFYVHFLDGSIFLVATGILSTAGQLP 416

QY 502 GATKRYGNQSEHTL-GVYHSHAFKVDLDVAGLENNVVAEDVAFYPMVPMSPERQL- 559

Db 417 GKNRPYGOSLNNDGLVAPIHQHMFNVRDFELDGVKNAVEVDMEY-----PEHNPT 468

QY 560 -QRLQVTRKLEMEQAAFLVGSATPRYLYLASNHS-NKMGHPRGYRIQLMSFAGEPLPQ 617

Db 469 GTAFMADRLLETQKARKNEAKHFRKATNIESKLVNVEPARYRL-----IPT 519

QY 618 NSSMARGSWERYQAVTORKEEPPSSSVFNQDMPAPVVD-----FSOFINNET-- 668

Db 520 NG-----IOLAA--RDDAVYSKRAQPARNNLWTVADRERFAAGEYPMQATGAD 567

QY 669 -----IAGKDLVAVNTAGFLIPIAEDIPNTVYNGNVGFLRYNFDDEPS 716

Db 568 DGLHWTQKDNINVDLVVYTFGMHVRLEDMV--VMPRONIGFLEPHGFENQPT 625

QY 717 FYSADSIYFRGDODAGAC 734

Db 626 LNLPTSTSTGTGADNC 643

RESULT 13

AMOI_ARTS1 STANDARD; PRT; 648 AA.

ID 007123;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Copper methylamine oxidase precursor (EC 1.4.3.6) (MAOXII).

GN MAOXII.

OS *Arthrobacter* sp. (strain P1).

OC Bacteria; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Micrococcales; Micrococcaceae; *Arthrobacter*.

OX NCBI_TaxID=47915;

RN [1]

RP MEDLINE=93374858; PubMed=8366046;

RX Zhang X., Fuller J.H., McIntire W.S.;

RT Cloning, sequencing, expression, and regulation of the structural

RT gene for the copper/topa quinone-containing methylamine oxidase from

RT *Arthrobacter* strain P1, a Gram-positive facultative methylotroph.;

RL J. Bacteriol. 175:5617-5627(1993).

CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

CC H(2)O(2).

CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinoxone per subunit.

CC -1- SUBUNIT: HOMODIMER.

CC -1- INDUCTION: BY METHYLAMINE.

CC autooxidation of a specific tyrosyl residue (by similarity).

CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAUQUINOXONE OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L12990; AAA22074.1; -
 DR HSSP: P46881; IAV4.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.
 DR Pfam: PF02728; Cu_amine_oxid3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 KW Oxidoreductase; Copper; TPQ; Metal-binding.
 FT PROPEP 1 9
 FT CHAIN 10 648 COPPER METHYLAMINE OXIDASE.
 FT MOD_RES 385 385 TOPAQUINONE (BY SIMILARITY).
 FT METAL 436 436 COPPER (POTENTIAL).
 FT METAL 438 438 COPPER (POTENTIAL).
 FT METAL 595 595 COPPER (POTENTIAL).
 SQ SEQUENCE 648 AA; 72805 MW; B2FB276749225385 CRC64;

Query Match 8.9%; Score 363; DB 1; Length 648;

Best Local Similarity 24.0%; Pred. No. 1.6e-20;
 Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

QY 55 GOSQLFADSLREELTVMRELTORLPGVDAQAARSDNCVSELTLPKAAALHLD 114
 DB 12 GSHPLDPLSRVETIARVALIKE--GPAAESR-----FISVELRESKDL----- 57
 QY 115 RGSPPAREALAIFFGROPQPNVSELVGLPRLPSYMDVYERRHGRLPYRRPVLQ 174
 DB 58 RAGVAAREADAVLV--DRAQARSFEAYVD-----LEAGTVDSWKLLENIGOPMLD 108
 QY 175 EYLDIQMIFNRELPOASGLHHCFCFKHGRNLVMTTPRGLQSGDRA---TWELLY 231
 DB 109 EPECEBAC--RKDPE-----VIAALAKRGITNDLVCFEPWSGYF 148
 QY 232 NISGAEFFLHVGLLELVNHALD-PARMTIQK--VFYQGRYDSLQLEAGEAGLVN 288
 DB 149 GEDEGRL--MRALFVVRREADSDPAHPIENPIFY-----DLNAG--KV 191
 QY 289 VLPDNGTGSWSLKSVPFGPAPLQFY-----PGPRFSVQSGRVASS 333
 DB 192 VRLDDQA-----IPVPSARGNYLPKYVGEARTDLKPLINTQPEASFTVGNHTWA 244
 QY 334 LMTFSGLAFSGPRIFDVRFGE-----RLVYEISLOEALAIYGNSPA-AMTRYVDG 388
 DB 245 DMSFRVGFTRPREGVLVQLKFKDQGVDRPINRASLEMYVPPGDTAPVQAKNAFSGE 304
 QY 389 FGAGKYTPPLTRGVDCPYLATYVDMHFLLESQAPKTRDAFCVFQONQGLPLRRHSDLY 448
 DB 305 YNIGNMANSLTLCGDGCEIKFYDGHSDVSHGNPTIEMNICHEEDST-----LW 356
 QY 449 SH--YEGGLAET-----VLVRSMSLTINYMDYTFHSGAIEIRYATGYISA-FLF 501
 DB 357 KHFDFREGTAETRSRKLIVSFATVANYEAFYWHLFJLDSIEFLVAKGILSTACQLP 416
 QY 502 GATGKGNQVSEHTL-GTVTHSHAFKVDLDVAGLEMMVAEDVFPMAVPMSPEQL- 559
 DB 417 GEKNPFGOSLNNDGLYPIQHMHFNMDLELDGVKNAYEVDEY-----PEHNPT 468
 QY 560 -QRLQYTRKLLMEQQAFLVGSATPRYLTLASNHS--NKMGHPRGYRIQMLSFAGEPLPQ 617
 DB 469 GTFAMVADRLLLETKAKIRKTNKAKHFKAIANESKNLVNPEAYVL-----IPT 519
 QY 618 NSSMARGFSWERYQLATORKKEEPPSSSVFNQNDPAAPYVD-----FSDFINNET-- 668
 DB 520 NG-----IQLAA--RDDAVYSKRAQFARNNLWTAVDTERFAAGEYPMQATGAD 567
 QY 669 -----IAGKLVANVTAGFLIHPAEDIPNTVYGVNGVGFRLRYNPFDEPS 716

DB 568 DGLHIWTKDRNIVDIDLVVWYTFGMHHVRLDWP--VMPRONIGFMLEPHGFNQNPT 625
 QY 717 FYSADSIYFRGDQDAGAC 734
 DB 626 LNPFTSTSTQTGTGEADTC 643

RESULT 14
 ID PAOX_ARTGO STANDARD: PRT; 638 AA.
 AC P46881;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phenylethylamine oxidase precursor (EC 1.4.3.6) (Amine oxidase).
 OS Arthrobacter globiformis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinobacteriales; Micrococcales; Micrococcaceae; Arthrobacter.
 OX NCBI_TaxID=1665;
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=IFO 12137 / ATCC 8010.
 RX MEDLINE=94197690; PubMed=8147851;
 RA Tanizawa K., Matsuzaki R., Shimizu E., Yoritani T., Fukui T.;
 RT Cloning and sequencing of phenylethylamine oxidase from Arthrobacter
 RT globiformis and implication of Tyr-382 as the precursor to its
 RT covalently bound quinone cofactor";
 RL Biochem. Biophys. Res. Commun. 199;1096-1102(1994).
 RN (2)
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 9-628.
 RX MEDLINE=98070236; PubMed=9405045;
 RA Wlce M.C., Doolley D.M., Freeman H.C., Guss J.M., Matsunami H.,
 RA McIntire W.S., Ruggiero C.E., Tanizawa K., Yamaguchi H.;
 RT Crystal structures of the copper-containing amine oxidase from
 RT Arthrobacter globiformis in the holo and apo forms: implications for
 RT the biogenesis of topaquinone";
 RL Biochemistry 36;16116-16133(1997).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- INDUCTION: BY PHENETHYLAMINE.
 CC -1- PTM: Topaquinone (TPQ) is generated by copper-dependent
 CC autoxidation of a specific tyrosyl residue.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U03517; AAA18114.1; -
 DR PDB: IAVK; 18-MAR-98.
 DR PDB: IAVL; 18-MAR-98.
 DR PDB: IAV4; 25-MAR-98.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.
 DR Pfam: PF02728; Cu_amine_oxid3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 KW Oxidoreductase; Copper; TPQ; 3D-structure; Metal-binding.
 FT PROPEP 1 2
 FT CHAIN 3 638
 FT MOD_RES 382 382 TOPAQUINONE.
 FT METAL 431 431 COPPER.
 FT METAL 433 433 COPPER.
 FT METAL 592 592 COPPER.
 FT DISULFID 317 343
 FT MOTAGEN 382 382

Y->F: LOSS OF ACTIVITY.

SQ SEQUENCE 638 AA; 70646 MW; 1800396BA7A93F2 CMC64;

Query Match 8.3%; Score 340; DB 1; Length 638;
Best Local Similarity 23.7%; Pred. No. 9.7e-19;
Matches 157; Conservative 100; Mismatches 274; Indels 132; Gaps 28;

OY 106 KAAALAHLD--RGSPPPREALATVFF---GROGQPNVSELVGPLRPHPSIMROYVER 159
 | | | | | :
DB 38 RAYIGVLDPARGASSEEDRRFRVFHIDVSCARQ-----EVTVSV 79
 | | | | | :
OY 160 HGGPL-----PYHRPVLFEQELDIDOMFN-----RETPOS 192
 | | | | | :
DB 80 TNGTVISAVELDTATAGTGLPELVEEFEEVQEQLATDERMLKALAARNLDVSKVRAPLSA 139
 | | | | | :
OY 193 GLHHCCYCKKHGRRLVMYMTAPRL----QGSDATNFGLYYNISGACFGLHHVGLELL 248
 | | | | | :
DB 140 GFEEFA---EEEGKRIL-----RELATVODEPDSAW---ANRPDGVAAYDVYSKEVT 187
 | | | | | :
OY 249 VNHKALDEARWITQKVYFOGRYYDSLADLAQFEAGLVVVVLIIPNGTGSWSLKSPVP 308
 | :
DB 188 --RVIDGVFPVPA--EHGNTPD-----ELGPFLT 215
 | :
OY 309 GPAPLQGYPOGPRRSVO-GSRVASSLTFFSGLCAFSGPRIFDVRFQ-GERL---VYEI 363
 | :
DB 216 TOKPRLSTIQPEGSPFTYVGNHIEMEKMSLDGVGFVREGVYLHNIAFRGDRLRPIINRA 275
 | :
OY 364 SLOELATLYGSGSPAAMTRYVD--GGFGMGKYTPLTRGVOC---PYLATVDMHFLLE 418
 | :
DB 276 SIENLVNYGDRSPRIKSNONYFTDIGELYGVGTANSLELGCDLGDITYLSPVISDAF--- 332
 | :
OY 419 SQAKPTINDAFCEVFQONOLPLRRHNSDLXHYFGSLAEIVLVKSMSTLLNYDYVMDTV 478
 | :
DB 333 -GNPREINCGICMHEEDGILAK--HSDLMGGINTRYRRNRBVVISFEFTIGNYDVGFWY 389
 | :
OY 479 FHPGSAIEIRRYATGYI--SSAPLFQATGKXGOVSEHNLGTYHTSANKFYDLVAGLEN 537
 | :
DB 390 LYLDSTIFFEAKATGVVTSAPREGSONT-SQLAPGLAPRHQHFASRLDMAIDGFTN 448
 | :
OY 538 WYMAEDMYFPMAYVWSEBEOHLQVTRKLEMEQQAFLVGSATPRXYLASNHS--NK 596
 | :
DB 449 RVEEDDVYRQMPCNGERNGANFSR---KRTVLTRESEAVREADANTGRMIWISNESKNR 505
 | :
OY 597 WGHPRGYI---QMSLFGCEPLPONSSMARGFSWERYQLATYQRKEEBSSSVF--NOND 652
 | :
DB 506 LNEPGYGLHAHNOFTLIADP---GSSIAARRAAFKTKMLWTARYXDDEYYPGDEVNONS 562
 | :
OY 653 PMAPTVDSDFINNNTIAGKDLVAVVTAAGFLIIIPAEDIPTNTVANGNGVGFLLRYPNFED 712
 | :
DB 563 GGAGLPST--IAODRDIDIGDIVVWHTGTGLHFRVEDMP--IMVVDVYGKFLRREGFED 618
 | :
OY 713 EDP 715
 | :
DB 619 RSP 621

RESULT 15
AMO_PEA ID AMO_PEA STANDARD; PRT; 674 AA.
AC Q43077;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
OS Amine oxidase [copper-containing] precursor (EC 1.4.3.6).
CS Pisum sativum (garden pea).
NC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxId=388;
RN (1)
RP SEQUENCE FROM N.A.
RA Tipping A.J., McPherson M.J.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
NN [2]

```

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC TISSUE-Seeding;
RX MEDLINE-96398683; PubMed-8805580;
RA Kumar V., Doolley D.M., Freeman H.C., Guss J.M., Harvey I.,
RT McGulir M.A., Milce M.C., Zubak V.M.;
RI "Crystal structure of a eukaryotic (pea seedling) copper-containing
RL amine oxidase at 2.2-A resolution.";
RU Structure 4:943-955(1996).
CC -I- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2);
CC -I- COPACATOR: Blinds 1 copper ion, 1 manganese ion and 1 topaquinone
CC per subunit.
CC -I- SUBUNIT: HOMODIMER.
CC CC -I- PMU: Topaquinone (TPQ) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (by similarity).
CC -I- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isd-slb.ch/announce/
CC or send an email to license@isd-slb.ch).
CC -----
DR EMBL, L39931; AAA62490.1; -.
DR PDB: 1KSI; 24-DEC-97.
DR InterPro; IPRO00269; CuNH_oxidase.
DR Pfam; PF011179; Cu_amine_oxid; 1.
DR Pfam; PF02727; Cu_amine_oxidn2; 1.
DR Pfam; PF02728; Cu_amine_oxidn3; 1.
DR PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE; PS01165; COPPER_AMINE_OXID_2; FALSE_NG.
KW Oxidoreductases; Metal-binding; Copper; Manganese; TPQ; Glycoprotein;
KV Signal; 3D-structure; 25.
FT SIGNAL 1..25 POTENTIAL.
FT METAL 26 674 AMINE OXIDASE [COPPER-CONTAINING].
FT METAL 476 476 MANGANESE.
FT METAL 477 477 MANGANESE (VIA CARBONYL OXYGEN).
FT METAL 478 478 MANGANESE.
FT METAL 617 617 MANGANESE.
FT METAL 618 618 MANGANESE (VIA CARBOXYL OXYGEN).
FT MOD_RES 412 412 TOPAQUINONE.
FT METAL 467 467 COPPER.
FT METAL 469 469 COPPER.
FT METAL 628 628 COPPER.
FT DISULFID 162 183 N-LINKED (GLCNAG. . . ) (PROBABLE).
FT DISULFID 344 370 N-LINKED (GLCNAG. . . ) (PROBABLE).
FT CARBOHYD 156 156 N-LINKED (GLCNAG. . . ) (PROBABLE).
FT CARBOHYD 389 389 N-LINKED (GLCNAG. . . ) (PROBABLE).
FT CARBOHYD 583 583 N-LINKED (GLCNAG. . . ).
SQ SEQUENCE 674 AA; 76358 MW; 30735390071DD18E CRC64;

Query Match      7.8% Score 320; DB 1; Length 674;
Best Local Similarity 25.8%; Pred.No.3.8e-17;
Matches 120; Conservative 66; Mismatches 189; Indels 90; Gaps 20.

QY 301 SLKSVPPEGPAPP--LQFYPPGGPRRSVOGSRYASLTMTFSFLGAFSG-----PRIFDV-   352
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 223 SKQSP-PGPKQHSLHSHQPGRPGIOHGHSVMANMKMFHFDFVRAGIVISLAISTYLE    291
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 353 RFQGRLLYEYSLOBALIIYGNSSPAAMTRYVDGC-FGMGYKTYPLRFGRVCDPYLATVV   411
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 292 KHKSRRVLYKGYSISLFAPYODPTFEFFFKFPFGSDGEFGSLSTVLPIPNRCPCPHAGFI   351
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 412 DMWHLLESQAAPTINDACVFEPONCGLP LRHHNLSLYIFEGGL-----AEVLVI     462
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 352 DTYYANSNGTPLLNNAICVEEQ-----YGNIMMRTENGIPNESIEESRTENLV       403
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 463 RSMSTLLNDYVMDTVTFPSGAIEIRFYATGISSAFLFGANGK-----YGVNSE        513
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 404 RTIVAVGYNVDYDIWFEPASGSKIKSFIALSIGLE---IKGTNIKKHDEIKEDLHGKIUSA    460
```

Tue May 20 13:18:55 2003

us-10-081-408-2.rsp

Page 15

```

QY 514 HTTITVTHSHNFYDQDVGLEWMAEDMVPFM-----AVPSPHQLQRLQVTR 566
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 NSIITIHDFHYITLDPDIDGTHSEFKTSILKTVIRIDGSSKRSYVTTETQAK---- 515
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 567 KILMEBOAFLVGSATPRYLILSNHNSNKGHRCGRIOMLSPAGEPL-----PONSS 620
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 ----TESDAKKTISLPAELVYVNPNIKTIVGNVGR-LIPAIPIHPLITLEDYRQ--- 567
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 621 MARGFSSBERIOLAVTKRKEEPPSSSVFNONDPAP--TVDFS---DFI-----NNETIA 670
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 568 -IRG-AFTNNVNVVT-----AYNRTKKAAGSLYUDHSQDITLAVMTKONREIV 614
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 671 GKQDVAVWTAGFLIHPAEDIPNTVTGNGVGFRLIRYFNEDEBP 715
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 NKQIVMMHHVGVIIHVPKQEDFPIPLL--STSELRPTNFFERNP 657
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: May 20, 2003, 11:34:58
Job time : 22.01 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 10:12:24 ; Search time 52.6446 Seconds

(without alignments)
2986.328 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095

Sequence: 1 MNOKTILVLLAVITTFAL.....QAACAPDLPAPFSGHGFSGHN 763

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*
15: SP-virus:*
16: SP-bacteriophage:*
17: SP-archaeo:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3517	85.9	763	6	Q9TK6 Bos taurus
2	3478	84.9	765	11	Q9R055 mus musculus
3	1404.5	34.3	751	11	Q8VC36 Q8VC36 mus musculus
4	1032	25.2	447	11	Q8R229 Q8R229 mus musculus
5	649	15.8	787	3	Q96X16 Q96X16 Pichia past
6	615.5	15.0	271	11	Q920K6 Q920K6 cavia porce
7	472.5	11.5	204	11	Q9CW16 Q9CW16 mus musculus
8	385	9.4	741	10	Q9C6V7 Q9C6V7 Arabidopsis
9	381	9.3	654	16	Q989X2 Q989X2 Arabidopsis
10	372.5	9.1	668	10	Q48552 Q48552 Arabidopsis
11	368.5	9.0	660	17	Q97XMI Q97XMI Arabidopsis
12	358	8.7	64	4	Q9UEU7 Q9UEU7 Homo sapien
13	357.5	8.7	756	10	Q9ST12 Q9ST12 Arabidopsis
14	357	8.7	194	4	Q9HBM7 Q9HBM7 Homo sapien
15	353.5	8.6	650	10	Q23349 Q23349 Arabidopsis
16	353	8.6	649	10	Q8W1C1 Q8W1C1 brassica ju

17	343.5	8.4	735	10	Q9SW88 Q9SW88 canavalia l
18	338	8.3	1794	10	Q9S168 Q9S168 Arabidopsis
19	335.5	8.2	422	10	Q9C6M1 Q9C6M1 Arabidopsis
20	328	8.0	674	10	Q9SXN5 Q9SXN5 Pisum sativ
21	327	8.0	687	10	Q9M2B9 Q9M2B9 Arabidopsis
22	320.5	7.8	670	10	Q65749 Q65749 Cicer ariet
23	319	7.8	420	10	Q9C6V9 Q9C6V9 Arabidopsis
24	318	7.8	701	10	Q9XHP4 Q9XHP4 Arabidopsis
25	310	7.6	677	10	Q9SW90 Q9SW90 glycine max
26	309	7.5	667	10	Q9LD03 Q9LD03 Lens culina
27	307.5	7.4	638	3	Q96X06 Q96X06 Aspergillus
28	302.5	7.4	759	10	Q9SLR1 Q9SLR1 Arabidopsis
29	300	7.3	660	16	Q8YRL5 Q8YRL5 Arabidopsis
30	279.5	6.8	712	3	Q9P7F2 Q9P7F2 Schizosacch
31	257.5	6.3	118	11	Q9D4M4 Q9D4M4 mus musculu
32	247	6.0	794	3	Q42890 Q42890 Schizosacch
33	246.5	6.0	300	10	Q9ST13 Q9ST13 Arabidopsis
34	200	4.9	103	11	Q63973 Q63973 Rattus norv
35	179.5	4.4	102	6	Q29317 Q29317 Sus scrofa
36	126	3.1	862	4	Q9WT23 Q9WT23 Homo sapien
37	125.5	3.1	460	10	Q9ST14 Q9ST14 Arabidopsis
38	125	3.1	757	16	Q9K6G6 Q9K6G6 Bacillus ha
39	115	2.8	2675	2	Q9FB23 Q9FB23 Streptomyce
40	114.5	2.8	949	5	Q952L5 Q952L5 Caenorhabd
41	113.5	2.8	1275	6	Q62658 Q62658 Canis fami
42	113	2.8	43	11	Q99N45 Q99N45 cavia porce
43	113	2.8	788	5	Q8STB6 Q8STB6 Drosophila
44	112.5	2.7	605	3	Q96V43 Q96V43 Aspergillus
45	112	2.7	470	17	Q8ZYN8 Q8ZYN8 Pyrobaculum

ALIGNMENTS

RESULT 1
ID Q9TK6 PRELIMINARY: PRT: 763 AA.

AC 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DI 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Semicarbazide-sensitive amine oxidase (EC 1.4.3.6).
GN SSO.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Iwabuchi H., Matsumura K., Mure M., Kuroda S., Tanizawa K.;
RT "Molecular cloning of semicarbazide-sensitive amine oxidase gene from
RT Bovine aorta.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB019242; BAA8896.1; -;
DR InterPro: IPR000269; Cunn_oxidase.
DR Pfam: PF01179; Cu_amine_oxid. 1.
DR Pfam: PF02727; Cu_amine_oxid. 2.
DR Pfam: PF02728; Cu_amine_oxid. 3.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID.1; UNKNOWN.1.
DR PROSITE: PS01165; COPPER_AMINE_OXID.2; 1.
KW Oxidoreductase.
SQ SEQUENCE 763 AA; 84500 MW; DB61ED9A89E71E90 CRC64;

Query Match 85.9%; Score 3517; DB 6; Length 763;
Best local similarity 85.2%; Pred. No. 3.2e-280;
Matches 650; Conservative 44; Mismatches 69; Indels 0; Gaps 0;

QY 1 MNOKTILVLLAVITTFALVCLVLRGGDGPSPQLPHCPSPVSPAPQWTHPGOSQLF 60
DB 1 MNOKTILVLLAVITTFALVCLVLRGGDGPSPQLPHCPSPVSPAPQWTHPGOSQLF 60

```
QY 61 ADLSREELTAVMRELTQRLGVLDAQAARSDNCVSEVLEQLPKAAALAHLDGSSPP 120
DB 61 ADLSREELTAVMSFLTKLGLDVLDAQAARSDNCISVLEQLPKAAALAHLDGSSPP 120
QY 121 AREALATVFEQROPQVNSVSELYGRLPHPSYMRDVTYERHGGPLPYRRRPLYFOEYLDID 180
DB 121 AREALATVFEQROPQVNSVSELYGRLPHPSYMRDVTYERHGGPLPYRRRPLYFOEYLDID 180
QY 121 AREALATVFEQROPQVNSVSELYGRLPHPSYMRDVTYERHGGPLPYRRRPLYFOEYLDID 180
DB 121 AREALATVFEQROPQVNSVSELYGRLPHPSYMRDVTYERHGGPLPYRRRPLYFOEYLDID 180
QY 181 QMIFNRELPOASGLLHHCCKYKRRGRNLVMTTAPRGLQSDRATWGLYINISGAFEL 240
DB 181 QMIFNRELPOASGLLHHCCKYKRRGRNLVMTTAPRGLQSDRATWGLYINISGAFEL 240
QY 241 HHVGLLELVNHNKALDPAKMTIQKRYGYRYSLAQLEAOEALVNVVLLPDNGTGGSW 300
DB 241 HHVGLLELVNHNKALDPAKMTIQKRYGYRYSLAQLEAOEALVNVVLLPDNGTGGSW 300
QY 301 SLKSVPPGPAFLQFPYPOGRFSVQGSRYVASSLMTFSFGAGSPRIPDRVROGERLY 360
DB 301 SLKSVPPGPAFLQFPYPOGRFSVQGSRYVASSLMTFSFGAGSPRIPDRVROGERLY 360
QY 361 YEISLOALAIYGNNSPAAMTRIVDGGFGNGKTTPLTRGVDCPYLATYVDHMFLESO 420
DB 361 YEISLOALAIYGNNSPAAMTRIVDGGFGNGKTTPLTRGVDCPYLATYVDHMFLESO 420
QY 421 APKTRDAFCVFEONOGPLRRHNSDLXSHYFGGLAETVLVVRSMSTLLNDYMDVYFH 480
DB 421 APKTRDAFCVFEONOGPLRRHNSDLXSHYFGGLAETVLVVRSMSTLLNDYMDVYFH 480
QY 481 PSGLAIEIRFATGYISSAFLEGGATGKYGNVSEHTLGTVHTSHAHFKVLDVAGLENNVW 540
DB 481 PSGLAIEIRFATGYISSAFLEGGATGKYGNVSEHTLGTVHTSHAHFKVLDVAGLENNVW 540
QY 541 AEDVAFVPMVPSPEHOLQRLQVTRKLLMEBOAFLVGSATPRYLYLASNHSKNGHP 600
DB 541 AEDVAFVPMVPSPEHOLQRLQVTRKLLMEBOAFLVGSATPRYLYLASNHSKNGHP 600
QY 601 RGYRIQMLSFAGEPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPYVDF 660
DB 601 RGYRIQMLSFAGEPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPYVDF 660
QY 661 SDFINNETIACKDLVAVWTAGFLIIPHAEDIPNVTVGNGVGFELRPVNFDEDEPSTYSA 720
DB 661 SDFINNETIACKDLVAVWTAGFLIIPHAEDIPNVTVGNGVGFELRPVNFDEDEPSTYSA 720
QY 721 DSIFRGDODGACEVNPPLACTPOAACAPDLPAFSGHGFESH 763
DB 721 DSIFRGDODGACEVNPPLACTPOAACAPDLPAFSGHGFESH 763

RESULT 2
QY 09R055 PRELIMINARY: PRT: 765 AA.
AC 09R055;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2002 (Tremblrel. 20, Last annotation update)
DE Copper amine oxidase (EC 1.4.3.6).
CN AOC3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99194802; PubMed=10092636;
RA Moldes M., Feve B., Paltrault J.;
RT "Molecular cloning of a major mRNA species in murine 3T3 adipocyte
RT lineage: differentiation-dependent expression, regulation, and
RT identification as semicarbazide-sensitive amine oxidase.";
RL J. Biol. Chem. 274:9513-9523(1999).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: COPPER AND TOPAQUINONE (BY SIMILARITY).
```

```
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
DR EMBL: AF115411; MAD09199.1; -.
DR MGD: MG1:1306797; AOC3.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid. 1.
DR Pfam: PF02727; Cu_amine_oxid2. 1.
DR Pfam: PF02728; Cu_amine_oxid3. 1.
DR PRINTS: PR00766; CUDROXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR KEGG: OXIDOREDUCTASE; TPO.
SQ SEQUENCE 765 AA; 84534 MW; 4252854759F316C7 CRC64;

Query Match 84.9%; Score 3478; DB 11; Length 765;
Best Local Similarity 83.38; Pred. No. 5,1e-277;
Matches 635; Conservative 60; Mismatches 67; Indels 0; Gaps 0;

QY 1 MNOKTILVLLIATVITFALVCVLLVGGGSGPSSOLPHCPSPVSPAOPWTHGQSOLF 60
DB 1 MNOKTILVLLIATVITFALVCVLLVGGGSGPSSOLPHCPSPVSPAOPWTHGQSOLF 60
QY 1 MTQKTTLVLAIAVITTFALVCVLLVGGGSGPSSOLPHCPSPVSPAOPWTHGQSOLF 60
DB 1 MTQKTTLVLAIAVITTFALVCVLLVGGGSGPSSOLPHCPSPVSPAOPWTHGQSOLF 60
QY 61 ADLSREELTAVMRELTQRLGVLDAQAARSDNCVSEVLEQLPKAAALAHLDGSSPP 120
DB 61 ADLSREELTAVMRELTQRLGVLDAQAARSDNCVSEVLEQLPKAAALAHLDGSSPP 120
QY 121 AREALATVFEQROPQVNSVSELYGRLPHPSYMRDVTYERHGGPLPYRRRPLYFOEYLDID 180
DB 121 AREALATVFEQROPQVNSVSELYGRLPHPSYMRDVTYERHGGPLPYRRRPLYFOEYLDID 180
QY 181 QMIFNRELPOASGLLHHCCKYKRRGRNLVMTTAPRGLQSDRATWGLYINISGAFEL 240
DB 181 QMIFNRELPOASGLLHHCCKYKRRGRNLVMTTAPRGLQSDRATWGLYINISGAFEL 240
QY 241 HHVGLLELVNHNKALDPAKMTIQKRYGYRYSLAQLEAOEALVNVVLLPDNGTGGSW 300
DB 241 HHVGLLELVNHNKALDPAKMTIQKRYGYRYSLAQLEAOEALVNVVLLPDNGTGGSW 300
QY 301 SLKSVPPGPAFLQFPYPOGRFSVQGSRYVASSLMTFSFGAGSPRIPDRVROGERLY 360
DB 301 SLKSVPPGPAFLQFPYPOGRFSVQGSRYVASSLMTFSFGAGSPRIPDRVROGERLY 360
QY 361 YEISLOALAIYGNNSPAAMTRIVDGGFGNGKTTPLTRGVDCPYLATYVDHMFLESO 420
DB 361 YEISLOALAIYGNNSPAAMTRIVDGGFGNGKTTPLTRGVDCPYLATYVDHMFLESO 420
QY 421 APKTRDAFCVFEONOGPLRRHNSDLXSHYFGGLAETVLVVRSMSTLLNDYMDVYFH 480
DB 421 APKTRDAFCVFEONOGPLRRHNSDLXSHYFGGLAETVLVVRSMSTLLNDYMDVYFH 480
QY 481 PSGLAIEIRFATGYISSAFLEGGATGKYGNVSEHTLGTVHTSHAHFKVLDVAGLENNVW 540
DB 481 PSGLAIEIRFATGYISSAFLEGGATGKYGNVSEHTLGTVHTSHAHFKVLDVAGLENNVW 540
QY 541 AEDVAFVPMVPSPEHOLQRLQVTRKLLMEBOAFLVGSATPRYLYLASNHSKNGHP 600
DB 541 AEDVAFVPMVPSPEHOLQRLQVTRKLLMEBOAFLVGSATPRYLYLASNHSKNGHP 600
QY 601 RGYRIQMLSFAGEPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPYVDF 660
DB 601 RGYRIQMLSFAGEPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPYVDF 660
QY 661 SDFINNETIACKDLVAVWTAGFLIIPHAEDIPNVTVGNGVGFELRPVNFDEDEPSTYSA 720
DB 661 SDFINNETIACKDLVAVWTAGFLIIPHAEDIPNVTVGNGVGFELRPVNFDEDEPSTYSA 720
QY 721 DSIFRGDODGACEVNPPLACTPOAACAPDLPAFSGHGFESH 762
DB 721 DSIFRGDODGACEVNPPLACTPOAACAPDLPAFSGHGFESH 762

RESULT 3
QY 08VC36 PRELIMINARY: PRT: 751 AA.
ID 08VC36
```


AC 08VC36:
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Similar to anilotide binding protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC021880; AAH21880.1;
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR000269; CUNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxid_2; 1.
 DR Pfam: PF02728; Cu_amine_oxid_3; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN.1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; UNKNOWN.1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; UNKNOWN.1.
 SQ SEQUENCE 751 AA; 85449 MW; 178B430795A2B988 CRC64;

Query Match 34.3%; Score 1404.5; DB 11; Length 751;
 Best Local Similarity 39.8%; Pred. No. 1.4e-106;
 Matches 298; Conservative 127; Mismatches 270; Indels 53; Gaps 15;

OY 45 SPSAQPWTHPGOSOLFADLSRELTAVMFLQRLGPGVDAQAQSPSCNCFVSVELQAP 104
 DB 19 TFSATVTHH-DARLFADISPOEIKAVHSFLMSRKELGLESSKNITLANSVFLIEMLP 77
 OY 105 PKAALAHLDHRSPPAREALAIFFGROPQPNVELVGPPLPHPSMDVVERHGP 164
 DB 78 KKNVTLKFLDEGRKSPVRARAIIFGAGDHNVTEFANGLPRCIYQALS-PRGHNL 136
 OY 165 PTHRRPVLEFOEYLDIDQMFNRELPOASGLHHCPCYKRG-----RNLVTMTAPR 216
 DB 137 SWSSPRISTAEY-DLLYHMLNRAITPL-----HOFFLDITGFSFLGCDREFLFTDVAPR 190
 OY 217 GLQSDRAVTFGLYNNISAGFELHVGLELLVNHKALDPAWITQKXVQRAYDLSAQ 276
 DB 191 GYESQGRSMILVQRYVE--GTFLEHTGLEILYDHSSTDVQDMKVEQLMCKFTNSPPE 248
 OY 277 LEAEGALVNVVLLPDNQTGGSWSLKSFPVPG-PAPPL-----QFY----- 317
 DB 249 LAQKAVGEVAVLEE-----VYLEDPLPGATEQPLFSSYKPRGEHHTVYAGPHV 302
 OY 318 --PQPRFSVQGSRYAASLMTFSFGLGAFSGPRIDVRFQGERLYVEISLOEALAYG 375
 DB 303 VQSPSPRKLKLEGNVLYLGDMSFSYRLRSSGLQIFNVLFGEGRVAYEVSVQEAVALYGC 362
 OY 376 SPAATTRVYDGFQMGKATYTPPLTRGVDCPYLATYVDHMFLEESQAPKTRIDAFVFEON 435
 DB 363 TPAGMGTIKYIDGKGLGSLTHLAPGIDCEPETAFLDAHYDSDGPLYPRALCLFEMP 422
 OY 436 OGLPLRRHSDLYS---HYFGGLAETVLVVRSMSTLLNDYVMDVTFHPSGALIEIRFYAT 492
 DB 423 TGVPLRRHSDLYS---HYFGGLAETVLVVRSMSTLLNDYVMDVTFHPSGALIEIRFYAT 492
 OY 433 GYISSAFLEGATGKXGNVSEHTLGTVTHSAHFQVDLDVAGLENNVMAEDMVFVMAVP 552
 DB 463 GYVAHATFTTPEGRLHGTQLTHLGNITHLVHVRVDLVAGTKNSFTLTKLENTINP 542
 OY 553 WSPHQRLQVYTRKLEMEEOAFLVGSATPRYLVLASNSNKGHPRGYRIOMLSFAGP 612
 DB 543 WPSHSLVQPTLEQTOYSHHQAAFRFGQTLPKYLLFSSPQKNRKHRSYRLQIHSMAL 602
 OY 613 EPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPVDFSDTI-NNETIAG 671
 DB 603 QVLPQMGQERAVVTAARYPLAVTKYRESERYSSSLYNQNDPMDPVPVVEEFLRNENIN 662

OY 672 KDLVAMTAGELHHPHAEDIPNTVVGNGEFLRPYNFEDEDPSEFSYSDSYFRGDODA 731
 DB 663 EDLVAMVTVGFLHHPHSEDPVNTATPGNCVGLIRPFNFEEEDPSLASHDVYI-V-PPQDN 721
 OY 732 GACEVNPPLACLPOAAACAPDLPAFSGHG 759
 DB 722 GLNHVQ--RWIPENRDCVLS-PPFSYNG 746

RESULT 4

OY 08R229 PRELIMINARY; PRT: 447 AA.

AC 08R229:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 51.3 kDa protein (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC022627; AAH22627.1;
 KW Hypothetical protein.
 FT NON TER
 SQ SEQUENCE 447 AA; 51265 MW; E556EAA2525D1D1 CRC64;

Query Match 25.2%; Score 1032; DB 11; Length 447;
 Best Local Similarity 44.6%; Pred. No. 2.8e-76;
 Matches 199; Conservative 78; Mismatches 161; Indels 8; Gaps 5;

OY 318 PQGPRFSVQGSRYAASLMTFSFGLGAFSGPRIDVRFQGERLYVEISLOEALAYG 377
 DB 1 PSGPRKLEGNVLYLGDMSFSYRLRSSGLQIFNVLFGEGRVAYEVSVQEAVALYGC 60
 OY 378 AAMTRVYDGFQMGKATYTPPLTRGVDCPYLATYVDHMFLEESQAPKTRIDAFVFEON 437
 DB 61 ACGMGTIKYIDGKGLGSLTHLAPGIDCEPETAFLDAHYDSDGPLYPRALCLFEMP 120
 OY 438 LPLRRHSDLYS---HYFGGLAETVLVVRSMSTLLNDYVMDVTFHPSGALIEIRFYAT 494
 DB 121 VPLRRHSDLYS---HYFGGLAETVLVVRSMSTLLNDYVMDVTFHPSGALIEIRFYAT 494
 OY 495 ISSAFLEGATGKXGNVSEHTLGTVTHSAHFQVDLDVAGLENNVMAEDMVFVMAVP 554
 DB 181 VHAFTFTPEGRLHGTQLTHLGNITHLVHVRVDLVAGTKNSFTLTKLENTINP 540
 OY 555 PPHQRLQVYTRKLEMEEOAFLVGSATPRYLVLASNSNKGHPRGYRIOMLSFAGP 614
 DB 241 PHSVSLVQPTLEQTOYSHHQAAFRFGQTLPKYLLFSSPQKNRKHRSYRLQIHSMAL 600
 OY 615 LPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPVDFSDTI-NNETIAG 673
 DB 301 LPPGMOEERAVVTAARYPLAVTKYRESERYSSSLYNQNDPMDPVPVVEEFLRNENIN 660
 OY 674 LVAMVTVGFLHHPHSEDPVNTATPGNCVGLIRPFNFEEEDPSLASHDVYI-V-PPQDN 721
 DB 361 LVAMVTVGFLHHPHSEDPVNTATPGNCVGLIRPFNFEEEDPSLASHDVYI-V-PPQDN 721
 OY 734 CEVNPPLACLPOAAACAPDLPAFSGHG 759
 DB 420 NHVQ--RWIPENRDCVLS-PPFSYNG 746

RESULT 5
 OY 096X16 PRELIMINARY; PRT: 787 AA.
 AC 096X16:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT	01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE	Lysyl oxidase.
GN	AOC1.
OS	Pichia pastoris (Yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; Saccharomycetaceae; Pichia.
OX	NCBI_TaxID=4922;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21131966; PubMed=11237259;
RA	Kucha J.A., Dooley D.M.;
RT	"Cloning, sequence analysis, and characterization of the 'lysyl
RT	oxidase' from Pichia pastoris."
RL	J. Inorg. Biochem. 83:193-204(2001).
DR	EMBL; AF358434; AAK49976.1; -
DR	InterPro; IPR000269; CUNH_oxidase.
DR	Pfam; PF01179; Cu_amine_oxid; 1
SO	SEQUENCE 787 AA; 89690 MW; D779EBF2BF766745 CnC64;
Query Match	15.8%; Score 649; DB 3; Length 787;
Best Local Similarity	27.3%; Pred. No. 2e-44;
Matches 201; Conservative 114; Mismatches 320; Indels 102; Gaps 21.	
QY	45 SPSAOPWTHPGOSQLFADLSREELTAVMRFLTQRLGPGVLVDAQAARPSDNCYFSVELDLP 104
DB	55 APTKNTW-----SLAKEEVEVDLHLSTY--NITEYTKADFESNYLVMTIELTP 103
QY	105 PRAAALHLNDRSPPREALATVFFGPGOPQNVSELVGLPHHSYMDVVERHNGPL 164
DB	104 NKTETLYLDEDGDLPRNARTVTFEGGESEYFEELKGLPL---VSDETIE---PL 156
QY	165 PYHRRPVLFQEYLLDDIDMIFNRELPOASGLHHCFCYKRGHNLVTMTAPRGL----- 218
DB	157 SFYNTNGSKSLPFEYGHLDRIKSAKSS-----FLNKNLNTIMRVLGELIGVPE 208
QY	219 -----QSDRATWTCGLY-----NISGAGFLVHVGLELVNHNKALDPA 257
DB	209 DMGCHSAAPQLDHPATGTATVDYGTGNTINTENDAEMLVPRGFEFK-----FDMTGRDVS 261
QY	258 RWTICKVFFQGYGYSLAQL-ENQEFAGLVNVVLLP-DNGTGSGMSLSPVPPG----- 310
DB	262 QMKMLEIYINNKVYYSAEELYEMOKDQDFVTLPKIDVDN--LDMTVIQRNDSAPIRHLD 318
QY	311 --APPLQFYPOGPRPSVQGSRYVASSL--WTFESFGLGAFSGPRIFDVRFOGERLVEISIQ 366
DB	319 DKSPRLVPEBERRRARAYDGEELYEFMMDMGFTSMRSRDGISFYDITTFGERIYVELSIQ 378
QY	367 EALAIYIGNSPAAMTTRYVDGFGMGKTYTTLTRGVDCYLATIYVDW-FLLESQAPKTI 425
DB	379 ELIAEGSDDPENQHTFYSDISYGVGNRPS-LVPGYDCPATGAGYFTTDFEYDEFYNRTL 437
QY	426 RQAFVCFEEDNOCLPLRRHNSDLXHYFGSLAETVLVYRSMSTLNTDYWDIVFHHSGAI 485
DB	438 --SYCFEENQEDYSLLRHNGASYSAI--TONPTLVNREISITGNDYNEFLKFFLDGTL 492
QY	486 EIRFATGYTSSAFLEFGAT-GKYGNOVSEHTLGTVTHFAHKVDDVDVAGLENNWMAEDM 544
DB	493 EYSVRAAGITQAGYNNPEISAPYGLKIHLYLSGSHFDHVLNKKVLDLDVGCTKRNASKIYM 552
QY	545 VEVPAVAVMSPHEQLOLQVTRKLEMEM-----QAALVGSATPRYLILASNHS 594
DB	553 KDVDVEYWPAPGTVYNTKQIAREVLEKEDFNGINMPDENGQILLIESA-----EEG 603
QY	595 NKMGRPRGRIOMLSFAGGPLPONS-SMAKGSWEYVQLAVQORKEEPPSSSVFQNDP 653
DB	604 NSFQGNRAVNIIPGGGVYIRKIVNSKSGPETONMAWSNLFTRKHQDEELRSSALTNTNL 663
QY	654 MAPTVDPSQFINNETIAGKDLVAMWTAGLGLPHADIDINTYTVGNGVGFELPRPYNEFD 713
DB	664 YDPPVNFNAFLDDESDGEDIVAVNVLGLHLPLNSNDLPTNTIFSTAHASFMLTPFNWFD 723
QY	714 DSEFYSADSIYFRGDOD 730

```

Db      724 ENSRDTQOVFTTYDDE 740
:::|:::|:
RESULT 6
ID Q920K6 PRELIMINARY; PRT: 271 AA.
AC Q920K6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DR 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DS Diamine oxidase (EC 1.4.3.6) (Fragment).
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
NC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidi; Cavia.
RX NCBI_TaxId=10141;
RN [1]
RP SEQUENCE FROM N.A.
ST STRAIN=HARTLEY; TISSUE=SMALL INTESTINE;
RA Kitanaka J., Kitanaka N., Takemura M.;
RT "Cloning of a cDNA for guinea-pig diamine oxidase (histaminase).";
RL Submitted (OCT-2001) to the EMBL/Genbank/DDBJ databases.
DR EMBL; AB073298; BAB70498.1; -
DR InterPro; IPR000269; CuNH_oxidase.
DR Pfam; PF01179; Cu_amine_oxid. 1.
DR PROSITE; PS01165; COPPER_AMINE_OXID_2; UNKNOWNW_1.
KW Oxidoreductase.
FT NON TER
SQ SEQUENCE 271 AA; 31323 MW; CEC46D64E4AA135D CRC64;

Query Match 15.0%; Score 615.5; DB 11; Length 271;
Best Local Similarity 49.1%; Pred. No. 2,3e-42;
Matches 115; Conservative 31; Mismatches 87; Indels 1; Gaps 1;

QY 491 ATGYSASFATCATKYGNQVSEHTLGTVHTSAPFKVDLDVAGLENNYMAEDMVFVPM 550
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1 AAGVGHATPTTPERGORYGRHLTHLLGNHTLHLRVDMDDVAGTNSFOTLOMKLENT 60
551 VWPSEPHQFORQVTRKLEMEQAFLVGSA TP RYLTLASNHSNKRWGHPRGYRIOMLSF 610
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB _61 NWMSGHHLVGTILQMSRSQERQAAPFRGCYLP LRYLTFSNKKRMWGHRGYRLQIISM 120
611 AGEPLPONSMMARGFSWERVQLAVTORKEEPSSSVFNQNDPMAPVDESDFI_NNETI 669
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 121 ABCVLPAGLPERRAATWARYPLAVTKYRESELSSSIYNQNPMPPPVYFOKELHNEDI 180
670 AKGLDVAMVTAFGLHIHPHAEIDIPNVTVYGNGVGFLLRPVNFEDEDPSPFSASDI 723
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 181 ENEDLVAMVTAGFLHIPHESDVPNPATPGNSVGFLLRPENFERKEDPSLASRDIV 234

RESULT 7
Q9CWL6 PRELIMINARY; PRT: 204 AA.
ID Q9CWL6
AC Q9CWL6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DR 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DS 1600012D06RIK protein (Fragment).
OS Abp1 OR 1600012D06RIK.
NC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
NC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
RX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
ST STRAIN=C57BL/6J; TISSUE=PLACENTA;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shiraawa A., Shibata K., Yoshino M., Itch M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
```

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Mikalido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gunningich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shihata Y., Storch K.F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuhl S.,
 RA Hayashizaki Y.,
 RA "Functional annotation of a full-length mouse cDNA collection.",
 RT Nature 409:685-690(2001).
 RL EMBL: AK005423; BAB2401.1; -.
 DR MGD: MGI:1923757; ADP1.
 DR InterPro: IPR000269; CUNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 DR NON_TER
 FT SEQUENCE 204 AA; 23861 MW; 49PR7CBD6CD0555C CRC64;
 SQ
 Query Match 11.5%; Score 472.5; DB 11; Length 204;
 Best Local Similarity 50.3%; Pred. No. 8.6e-31;
 Matches 95; Conservative 24; Mismatches 65; Indels 5; Gaps 4;
 Oy 572 EQQAAFLVGSATPRILYLASHNSKMGHPGRYRIQMLSFAGEPLPONSSMANGFSWERYO 631
 Db 15 EHQAAFRFGQTLPLKYLTFSSPOKMRGHRYSRLTIHSAEOLVPLPGMOEBEAVTWARYP 74
 Oy 632 LAVTARKKEPSSSSVFENNDMPAPTVDESEFI--NNETAGDLYAVMTAGFLHPIHAD 690
 Db 75 LAVITKRESEYSSSLTINNDMPDPVVEEFLRNMENTENEDLVAMVTGVLHHSDD 134
 Oy 691 IPIITVYGVNGVGFLLRPYNFDEDEPSFYSDSIYFRGDODAGACEVNPACLPQAAACAP 750
 Db 135 VENTATPGNCVGFLLRPFNFEEDEPFLASRDYIV--WPDNGLNHYQ--RWIPENDCLV 191
 Oy 751 DLPAFSGG 759
 Db 192 S--PPFSYNG 199
 RESULT 8
 Q9C6V7 PRELIMINARY; PRT; 741 AA.
 ID Q9C6V7
 AC Q9C6V7
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Copper amine oxidase (EC 1.4.3.6).
 GN F2/H3_13.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen R., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hulzer L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.O., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz G.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu X.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
 RA Miltner J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Tottum M.J., Town C.D.,
 RA Ullrichbeck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana".
 RT Nature 408:816-820(2000).
 RL "CATALYTIC ACTIVITY: KCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
 H(2)O(2)."
 CC -1- COFACTOR: COPPER AND TOPAQUINONE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 DR EMBL: AC074360; AAG60154.1; -.
 DR HSSP: 043077; 1KST.
 DR InterPro: IPR000269; CUNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR Pfam: PF02727; Cu_amine_oxidn2; 1.
 DR Pfam: PF02728; Cu_amine_oxidn3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 DR COPPER_Oxidoreductase, TPO.
 KW SEQUENCE 741 AA; 84279 MW; DB69E786864B558 CRC64;
 SQ
 Query Match 9.4%; Score 385; DB 10; Length 741;
 Best Local Similarity 25.3%; Pred. No. 9.2e-23;
 Matches 183; Conservative 97; Mismatches 280; Indels 164; Gaps 33;
 Oy 50 PHTPGQSQLFADLSHEELTAVRFLTORLGRGLVD-----AAQAPSDNCVFS----- 98
 Db 104 PPPHP-----FPLTEILEINLVNINERYPIGLEHRTFOYGLNEPKSLVSSVSSQ 158
 Oy 99 -VELQPPKAAALAHDRSPPARALAVFGRGPQGNVSLV--VGLPHPSYKRDVT 156
 Db 159 YHNKSPPPQAFVIARDHK--TRE--IVDFA--SOAIVSKIHVG----- 199
 Oy 157 VERHGGPLHYHRPVLFOEYLDIDQMIFNREL-----PASGLHHCCEYKRGRL--V 209
 Db 200 ---NGYPM-----LTIDEQATSELTAKKPRPDSI-----RRGLNVSEV 237
 Oy 210 TMTTAPRGLOSGDRATWFG-----LYNISGAGFLAHVGLVNHKRLDPA 257
 Db 238 VVTSITMG-----WFEAKPERLIKRRPFYLNCSVMTYLPIDGMITIV--LDQM 286
 Oy 258 RMTIQKVFYQGRYDSLQLEAOFEGALNVNVLIPNGTGSMS--LKSPPPEPAPPLQF 316
 Db 287 KVT-----FRDRFTSPLPVA-----KTEFRISKLPKPGPSLQNAVLF 326
 Oy 317 YPOGPRFSYQGRVAVASSLMTF--SFGIGA--FSGPRIFDVROGER--LYEISLOEALA 370
 Db 327 QSGPPEFKIDGHNRRANNEFHMSFDVRAGLVSLASIFDMDNKRVOLYKHLSEIV 386
 Oy 371 YTGNSPAAMTTRYVD--GGFGMGKYYTTPLTRGYDCYLAITYDMHFLLESQAOKTRDAF 429
 Db 387 PYMDPSEMDYFRFFFCGCGEQYAVSLEPYDCGNMAFMGVAFSADGCPPIKITTNW 446
 Oy 430 CYFEONQGLPLRNHSDLSHYFGCLAETVLYVRSMTLLNVOYWDVTFHPGSAIEIR 489
 Db 447 CIFEKTAG--DIMKRHTEIEIPGLKVRPDSLVLRAMVTYVGNVDYIVDEYKPGSGIKIGV 505
 Oy 490 YATG-----YISAFLEGATKXYGNQVSEHTLCTVHTSHAFVVDLDVAGLEMYWA 541
 Db 506 GLTGVLVEVPKVKYNTSEI--KEDDHGITYADNTIGVNDHFTYTRLLDDIDOSTDSFVR 564
 Oy 542 EDWTF--VPMAYVWSEHOLQRLQVTRKLIEMEQAAFLVGSATPRILYLASHNSKMGH 599
 Db 565 SELVTRKTPKSV-----NPKRSYMTTKRLKAE-----LVVNPSPKRYKHGN 607
 Oy 600 PGCYR--IQMLSFAGEPLPONSSMANGFSWERYOLAVTORKKEPSSSVF--NQMDP-- 653
 Db 608 EYGRLLHGPASEGPLADDYPOIRAFTNTNVTTPYNTNTEWVMSGLYADRSQDDDTL 667
 Oy 654 --WAPTVDESDFIINNETINGKDLVAVMTAGFLHPIHADIPNTVTVGVNGVGFLLRPYNNF 711

408 LLNDYVMBIVFHPSSAIEIRRYATGI-SSAFLGATGKXGNQVSEHTLGTVHTSHAHF 526

QY 148 --HPSYMDVYVERHGGLPRHRYVLFQEYILDIDOMIFNRELPOASGLHHCCTYKRG 205
 DB 134 HGPFSEFTELEFKASKLTLTY--PPFKSLID-----RS 165
 QY 206 RNLYMTAPRGLQSGDRATWFG-----LYYNISGAGFLHHV-GLELLVNHKA 253
 DB 166 LNISVSCIP-----FTVGKGETTTRRELKASCFYDGSVNTTRPEIGITITID--- 216
 QY 254 LDPARMTIQKVFYQGRYDLSLAQLEAOFEAGLVNVLIPDNGTGSWSLK-SPVPPGAP 312
 DB 217 VD---SMQVVKYSDFRKKPLPEKE-----GNDFRKKHKPF----- 248
 QY 313 PLQFYPOGRPSVQSGRASSLSMTFSFGLAESSGRIFDY-----RFGQERLYEIS 364
 DB 249 PPSGVSTGRTIGNRKRNKAMKHFVGTAAAGVTISTASVLDRTKRFR--RVMYRQH 306
 QY 365 LQELALATGNSSPAAMTTRYVD-GGFGMGKTYTPTLRGVDCPYLATYVDMHFLLESQAPK 423
 DB 307 VSETFVPMDPTEMYETTFMDIGFGRSAVNLQPLDQONAFDGHVAGSDGTNQ 366
 QY 424 TIRDAFCVFEONQGLPLRRHSDLY---SHYFGLAETVLVVRNSMTLLNDYVMDTVFH 480
 DB 367 KMTNMCVFEKN-GYGASFRHTEINVPQVITSGAAEISLVRYMVAFLGNVDYIVDWEK 425
 QY 481 PSQATEI-----RFATGYSASFLEGATGKNGOVSEHTLCTVTHSHAFKVDY 532
 DB 426 KSGALRVVDLTGLVEKATSTYSNEQT--RENVTGTOVAKTIVNHDHYLYYLDLV 483
 QY 533 AGLENMVAEDMVEFVMA--VPMSPHQLRLQVTRKLEMEQOAFVLSATPRLYLAS 591
 DB 484 DQNGSLVAKAKTKYRVADVNTKSRSKSYWTVYKETAETDQVRLGSEVEYLLIYNP 543
 QY 552 NNSKMGHPRGYRIQMLSFAGEPLPONS-----SMARQSWERYOLAVTQKEEBS 643
 DB 544 NKKTIGTGVYRL-----IPEHLPATSLTDDYETIAGYT--KYPWMT----- 588
 QY 644 SSSVNONDPNAPRTYDPSD-----FINNETIAGKDIYAVATGFLIIPHAEDIP 692
 DB 569 ---AYDKSERMAGGF-YSDRSRGDGLAVWSRNRLEIKNDIWMYNGFHIIPOEDIP 644
 QY 693 NTVTGNGVGFELRPYNFEDEDP 715
 DB 645 VMPRLHG--GFTLRPSNFEEDNP 665
 RESULT 11
 Q97XMI PRELIMINARY; PRT; 660 AA.
 ID Q97XMI AC Q97XMI
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Amine oxidase (copper-containing) (tyna) (EC 1.4.3.6).
 GN TYNA OR SSO1692.
 OS Sulfolobus solfataricus.
 CC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCB1_taxid=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35093 / DSM 1617 / P2.
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aveyez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erasuo G., Fletcher C., Gordon P.N.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,
 RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 RL EMBL; AE006782; AAK41902.1;
 DR InterPro; IPR000269; Cunn_Oxidase.

DR Pfam; PF01179; Cu_amine_oxid. 1.
 DR Pfam; PF02727; Cu_amine_oxid2; 1.
 DR Pfam; PF02728; Cu_amine_oxid3; 1.
 DR PROSITE; PS01164; COPPER AMINE OXID. 1; UNKNOWN. 1.
 DR PROSITE; PS01165; COPPER AMINE OXID. 2; UNKNOWN. 1.
 KW Oxidoreductase; Complete proteome
 SQ SEQUENCE 660 AA; 77136 MW; 413121D91C74076 CRC64;
 Query Match 9.0%; Score 368.5; DB 17; Length 660;
 Best Local Similarity 24.4%; Pred. No. 1.8e-21;
 Matches 175; Conservative 103; Mismatches 303; Indels 137; Gaps 28;
 QY 63 LSRELTAVMRLTQRLRGLVDAQAQAPSDNCVSEVLQLPKAAALAHLDGSPPPAR 122
 DB 17 LNEEIRKRSVEELKROLN---LDAKVVK-----PFSVELKEKQOYELMRNNKNIKIER 68
 QY 123 EALATVFGROPQPNVSELVVGRLPHPSYMRDVTVERHGGPLPYRRPVLQGEYLDIDOM 182
 DB 69 ESL-VKTYNFDQR-KYTEAIT-----SLDNNVKEIRSIDAY--PRTTDEFECEKA 118
 QY 183 IFNRELPOASGLHHCCTYKRGRLVMTTAPRGLQSGDRATWFGLYYNISGAGFLHH 242
 DB 119 VANDKRVOEA-----LTKRGILINDNL--LWVDCWAPG---H 151
 QY 243 VGEILLVNHKALDPAKMTIQKVFYQGRYDLSLAQLEAOFEAGLVNVLIPDNGTGSWSL 302
 DB 152 VDEELRGRRVVAIG-YMW-VKNDIEDNGCRPHGLMPWVDLKMVEIRIDHDGT----- 203
 QY 303 KSPVPPGPA-----PPLQF-YPGQPRFSVQSGRVASLSLTFSEFGAFSG 346
 DB 204 -SPLQLQDANITPEKLGITFGDDLPIETROPLOSSIKINGEISITRRRLIGITPRRG 262
 QY 347 PRIFDVRQGER---LYTEISLQELALATGNSSPAAMTTRYVDG--FGMGKTYTPTLR 400
 DB 263 LVIYDVRYIDENNRREMLILRASVYVDLWVYGDPSFHNKKMYLADGVDGLGNFTPLPSH 322
 QY 401 G-----VDC-PYLATYVDMHFLLESQAPKTRDAFCVEONQGLPLRRHSDLYSHYG 453
 DB 323 GWYDLYNDCDFEGEVYIHLDTVRSSNGTPIKKKAICVAHEDFGVLWR--HIDLRS--- 376
 QY 454 GLAET---VLVVRNSMTLLNDYVMDTVFHRSGAIEIFYATGYISSAFLF--CATGK 507
 DB 377 GKSEYRRNRRLVVSFMATLANTDYGFYFYFDGSEFLYKLTGITINDSISEKPTPKY 436
 QY 508 GNQVSEHTLGTVTHSAHEKVDLDVAGLENMVAEDMVEFVMAVPMSPHQLRLQVTRK 567
 DB 437 GTRVTPVYAPRIHIFENIRLNIINVDGLNRIT-----EVNLRGEPIERK 481
 QY 568 -----LLEMEQOAFVLSATPRYLYLAS-NNSKMGHPRGYRIQMLSFAGEP 614
 DB 482 NPGVNAFFAEENLLENEADARRHVPQIGRYKAIYNIQKNLGLPVARLVPCHNVLP 541
 QY 615 LPONSSMARGFSEMEROLAVTQKEEPPSSSVFQNDMPATVPSD-----FINNETI 669
 DB 542 LPDDSVRRRGAYINTHLWVTPYNEERYASDY-----PYLRANDLPYTIILKNSI 594
 QY 670 ACKDLVAMVTAGFLIIPHAEDIPNTVTGNGVGFELRPYNFEDEPSEFSADSIYFRG 727
 DB 595 VDEDLVITWTLGVEHVRIEDMP--VMPVEMAGFRILPQGFIDKNTIYLPQRLRING 650
 RESULT 12
 Q9UEU7 PRELIMINARY; PRT; 64 AA.
 ID Q9UEU7 AC Q9UEU7
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Placenta copper monamine oxidase (Fragment).
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCB1_taxid=9606;

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-96207227; PubMed-8619474;
RA Anderson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
RT "A 'double adaptor' method for improved shotgun library
construction";
RL Anal. Biochem. 236:107-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-97264341; PubMed-9110174;
RA Yu W., Anderson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
RT "Large-scale concatenation cDNA sequencing";
RL Genome Res. 7:353-358(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Yu W., Gibbs R.A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF034985; AAC09346.1;
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
FT NON_TER
SQ SEQUENCE 64 AA; 6892 MW; A7A0F2972EAF3210 CRC64;
Query Match 8.7%; Score 358; DB 4; Length 64;
Best Local Similarity 100.0%; Pred. NO. 3.9e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 700 GGGFLLRPNFEDDESFYSADSIYRGDODAGACEVNPLACPOAAACAPDLPAPSHG 759
DB 1 GGGFLLRPNFEDDESFYSADSIYRGDODAGACEVNPLACPOAAACAPDLPAPSHG 60
QY 760 FSHN 763
DB 61 FSHN 64
RESULT 13
Q9ST12 PRELIMINARY; PRT; 756 AA.
AC 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
GN 14C9.130 OR AT4G12290.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID-3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
BA Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Mannhaupt G.,
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Mewes H.W.,
BA Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL, AL080318; CAB45976.1;
DR EMBL, AL161533; CAB78272.1;
DR HSSP; Q43077; IKSI.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; UNKNOWN.1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
SQ SEQUENCE 756 AA; 84910 MW; 891019DEF0F649CA CRC64;
Query Match 8.7%; Score 357.5; DB 10; Length 756;
Best Local Similarity 22.2%; Pred. NO. 1.7e-20;
Matches 163; Conservative 100; Mismatches 280; Indels 191; Gaps 29;
QY 94 NCVFSYELQPPKAAALAHLDGSPPARALAI-----VEFGROPQPNVSE 140
DB 126 HALHETVLEPERK-NLYRHEKGNPLPRKASVIARVAGDTYLVLDISTGRVDSNSPV 184
QY 141 LVVG-PLPHPSYMRDVTYRHHGRLPYHRRPVLFOGYLIDIDIMIFRELPOASGLLHNC 199
DB 185 RVSGIPMTITEKNDITV-----VFSNAD-----FKRTIIS----- 216
QY 200 FYKHGRNLVTMTTAPRGLOSGDRATWFG-----LYNISGAGFLHHV-GLE 246
DB 217 ---RGVNLTDVIGCP--ISCG---WFGNKENARVIKSGCFMGTGTFNFMPIEGLT 266
QY 247 LLYNKAALDPARTIOKVYQGRYDLSLAOLEAFAGLVNVLIPDNGTGSWSLSKSPV 306
DB 267 ILID--LDTKQ-----PAPLQF-YPGGRFSVGSRYAS-SLMFSPGLAFSP 347
QY 307 PEG-----PAPLQF-YPGGRFSVGSRYAS-SLMFSPGLAFSP 347
DB 289 PGSTNDYRFOKLATTDKRPLNPISIEOPRGSPSEIENHLKMANWEHLKPDPRAY 348
QY 348 RIFDVR-----FOGRLVYEISLOEALAIYGSNSPAAATTRYVDG-FGNGKYTTLTR 400
DB 349 VISRVAVHDPDHEHEDVNYKGVSELFVPMYDSDAWYFKTMDAGEYFGGLQAMPILVP 408
QY 401 GVDCLPLATYVDMHFLLESQAPKTIIDACVPEQN-----OGLPLRRHSD 447
DB 409 LNDGPRNAAVMDGVFAADGTPEVRNMCIFESYAGDIGRHSSEPIGIPREVR--- 465
QY 448 YSHFGLAETVLYVMSMTLLNYDVMPTVFPSPGATIEIRFYATGYSASFGLATGY 507
DB 466 -----PKVTLVYMAASVGNVDIIDIERYOTDGLIAKGLSGL---MKGTYGN 514
QY 508 GNOV-----SEHTLGTVHTSAHFVDDVAGLENNVMAEDMVEFPMAY 551
DB 515 KNOVERKDQONEELHGLTLLSERVIGVINDHYVTFLDLDVDPDN-----SPVKYNLR 569
QY 552 PMSPEHQLOR---LOYTRKLEMEQDAFLVGSATP-RILYILASNSNKGHRGRIRIOM 607
DB 570 QETEPESPRKSYLKAVRNIAKTEKDQIKLSLYDSEFHHVINSCTTRVGNTPGKVPV 629
QY 608 LSPAGEPLPONSSMARGFSWERYOLAVTORKEEPESSSVFNONDMAPATVD-FSPFINN 666
DB 630 RTTAAALILHDHDPQKRGAFNTQNIWTPYKSEQAGLFTYQSGDDTLIATWSB--RD 687
QY 667 ETIAGDLVAVMTAGFLIHPHAEDIPNTYVGVNGVFLRPYNFEDDESFYSADSIYR 726
DB 688 ROIENKDIYVWYTLGFHHIFPCQEDFPIPTVSS--SFDLKPVNFFERNPLTSAAPNF--- 742
QY 727 GDDDAGACEVNPLA 740
DB 743 -EHDLPVCGVSVS 755
RESULT 14
Q9HBM7 PRELIMINARY; PRT; 194 AA.
AC 09HBM7;
DT 01-MAR-2001 (TREMblrel. 16, Created)

01-MAR-2001 (Tremblrel. 16, last sequence update)
01-MAR-2002 (Tremblrel. 20, last annotation update)
Hypothetical 23.0 kDa protein.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.O., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Nevel Human cDNA clones with function of inhibiting cancer cell
RT growth";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF218035; AMG17277.1;
DR InterPro: IPR000269; Cu_amine_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
KW Hypothetical protein.
SEQUENCE 194 AA; 23024 MW; 5C486671AE16A149 CRC64;
Query Match 8.7%; Score 357; DB 4; Length 194;
Best Local Similarity 38.0%; Pred. No. 2.5e-21;
Matches 68; Conservative 31; Mismatches 80; Indels 0; Gaps 0;
QY 485 IEIRYATGYSSAFLFCATGKNGVSEHTLGTVHTSAHKYVDLVAGLENNYWAEDM 544
DB 1 MEAKHAAGYVATYTPBGLTRHLHGNTHLVHRYDLVAVGTNSFOTLDM 60
QY 545 VEPVAVWSPBHQLRQVTRKLEMEQAFVLSATPRVLYLASNSKMGPRGR 604
DB 61 KLENTNMRSPRHRYVQTLTEOTYSWEROAARFRKRLPKLLTSPQENMGKRSR 120
QY 605 IOMLSFACEPLPONSMMARGSWERYQLAVTORKEEPPSSSVFNQNDPAVDFSD 663
DB 121 LOHSMADQVLPQWQEQALITARYPLAVTKYRESELCSSIIYQNDMPHPSLSS 179
RESULT 15
ID 023349 PRELIMINARY; PRT; 650 AA.
AC 023349;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Copper amine oxidase (EC 1.4.3.6).
GN Atg14940.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=9612113; PubMed=9461215;
RA Beyan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambolt R., Wetzenecker T., Pohl T.M., Terryn N.,
RA Giesen J., Villalobos R., De Clerck R., van Montagu M., Lecharny A.,
RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA Entian K.D., Rieger M., Schaeffer M., Pons A., Puidomenech P., Doua A.,
RA Sliver M., James R., Montfort A., Pons A., Puidomenech P., Doua A.,
RA Vouklatou E., Milioni D., Hatzopoulos P., Pitarandi E., Obermayer B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rechman S., Ansoorge W., Cooke R., Berger C.,
RA Delseny M., Voet M., Volckaert G., Mewes H.W., Kloterman S.,
RA Schueller C., Chaltatzis N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana";
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COPFACTOR: COPPER AND TOPAQUINONE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
DR EMBL: 297337; CAB10273.1;
DR EMBL: AL161540; CAB78536.1;
DR HSEB: Q43077; IKST.
DR InterPro: IPR000269; Cu_amine_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
KW Copper; oxidoreductase; TPO.
SEQUENCE 650 AA; 73783 MW; 28741CEB3A43FF5D CRC64;
Query Match 8.6%; Score 353.5; DB 10; Length 650;
Best Local Similarity 23.9%; Pred. No. 2.9e-20;
Matches 161; Conservative 96; Mismatches 278; Indels 139; Gaps 29;
QY 99 VELOLPPKAAALAHL---DRGSPPARALAIFFGRQPPNVSELVVGRPLPHSYMRD 154
DB 56 LDLEPNKSHVQLWSPNPSKKPPRRRSFVVRAGG---TYELLT-----D 101
QY 155 VTVER-----HGP---LPYHRRVLFQETLDIDQIMFNNEL--PQASGLIHCC 199
DB 102 LITSIASSRIYTGQFPSTFTELFKASKLPLTPPFKSKILDRSLNISEVCIPTFG 161
QY 200 FYKHGRNLVMTTAPRGOSGDRAATWFLYLYNISAGFLNHV-GLELVNKKALDPAR 258
DB 162 WIGE-----TTTRELKA-----SCFRDGSVNVFTPIESITVTD---VD--- 200
QY 259 WTIQKVFYGRYDLSLAOLEAQEAGLVNVVLIIPDNGTGSWSLSPVPGAPPLQFY- 317
DB 201 -SMQVTKYSDRRKP-----IPDK-EGNDPRTK-----RPPFFC 234
QY 318 -PQGRFVSQSRVASSLTFSGLGAFSGPRFDV-----RQGERLYYEISLQRA 368
DB 235 NVSDTGFKLLGNRVKAMANKFHVGTARAGVITSTASVLDPRTKRR--RVNRGVHSET 292
QY 369 LAIYGNSPAAATRTYVD-GGFGMGKVTTPPLTRGVDCPYLATYVDVHFLLESQAPRTIR 427
DB 293 FVPYMDPTIEMYYRTFMIDGEGFGSAVNDPLIDCPQMAAFLDGHVAGPDTAQKTN 352
QY 428 AECVEQONQDPLRRHSDLY--SHYFGLAETVLVVRMSSTLLNTDYVMDTVFHPGA 484
DB 353 VMCVEFEKN-CYGASFRHTEINVPQVITSGEAEISLVVAVATLGVYDIDVMEFKNGA 411
QY 485 IEL-----KFATGYISSAFLFCATGKNGVSEHTLGTVHTSAHKYVDLVAGLE 536
DB 412 IRVGVLDGLVLEVKATSYNSDOI--TENYVGTILVAKNTIANNHDIYLLYYLDLDVNG 469
QY 537 NNWMAEDWVFPMA-VPAPEHQLQVTRKLEMEQAFVLSATPRVLYLASNSHN 595
DB 470 NDLVAKLKTIVATVENKSSRRKSYWYVKEAKTEADGRVRLGSDPELLIVNPKKT 529
QY 596 KMGHPGRYIOMLSFAGEPLP--ONSSMARGSWERYQLAVTORKEEPPSSSVFNQND 652
DB 530 KIGNTVGYRL-----IPEHLQATSLTDDYDELNAGYT---KPVVWTVADRSE 576
QY 653 PPAPTYDESD-----FINNTTAKGLDVANVTAGFLHPIAEDIPNTVYGNV 701
DB 577 RWAAGG-YSDRSGDDGLAVWSSRNREIKDILVMYVNGVFHPIYQEDFPVMPPTLGG-- 633
QY 702 GEFLLRPYNFFEDDP 715
DB 634 GFTLRPSNFNDP 647

Search completed: May 20, 2003, 11:37:44
Job time : 56.6446 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 84.2615 Seconds

(without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357
Sequence: 1 MDMLRLFLMAAASINNA.....QAACAPDLPAFSGHGFSSN 998

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3967	74.1	763	20	AAV03219 Amino acid sequence
2	2480.5	46.3	729	23	AAU084261 Human endometrial
3	1180	22.0	447	23	AAU10670 Mouse L1CAM F4-5
4	1175.5	21.9	644	23	AAU10652 Chicken Ng-CAM F80
5	1175	21.9	536	20	AAV21557 GS-GC6 fusion prot
6	1171.5	21.9	514	19	AAW47004 Glutathione-S-tran
7	1168	21.8	1252	22	AAE02008 Yeast cystathionin
8	1167.5	21.8	362	16	AAV79948 G17 antigen. Chlm
9	1166	21.8	658	21	AAV96056 Human suppressor o
10	1164.5	21.7	443	23	AAU10673 Chicken Nr-CAM Fm4

11	1164	21.7	285	13	AAV22141
12	1164	21.7	442	23	AAU10667
13	1163	21.7	1196	21	AAV58917
14	1163	21.7	1196	21	AAV68853
15	1162.5	21.7	409	23	AAO17115
16	1162.5	21.7	409	23	AAW49711
17	1162	21.7	315	13	AAV22138
18	1162	21.7	361	13	AAV22145
19	1161	21.7	272	19	AAV56327
20	1161	21.7	635	23	AAU10668
21	1160.5	21.7	258	20	AAV74117
22	1160.5	21.7	277	20	AAV74116
23	1160.5	21.7	298	23	AAV87086
24	1159.5	21.6	259	20	AAV74125
25	1159.5	21.6	277	20	AAV74099
26	1159.5	21.6	277	20	AAV74107
27	1159.5	21.6	279	20	AAV74097
28	1158.5	21.6	426	16	AAV87091
29	1158	21.6	282	20	AAV74123
30	1158	21.6	379	20	AAV31954
31	1158	21.6	439	23	AAU10664
32	1157.5	21.6	282	20	AAV74095
33	1157.5	21.6	422	20	AAV23648
34	1157.5	21.6	559	23	AAU10663
35	1157.5	21.6	647	20	AAV73573
36	1157	21.6	247	20	AAV74114
37	1157	21.6	257	20	AAV74124
38	1157	21.6	516	19	AAV47005
39	1157	21.6	564	16	AAV84565
40	1157	21.6	643	16	AAV84568
41	1157	21.6	673	17	AAV98343
42	1156.5	21.6	247	20	AAV74103
43	1155.5	21.6	1140	16	AAV81331
44	1155.5	21.6	1140	23	AAE13360
45	1155	21.6	259	20	AAV74130

ALIGNMENTS

RESULT 1	AAV03219	standard: Protein; 763 AA.
ID	AAV03219:	
XX		
AC	21-JUN-1999	(first entry)
XX		
DE	Amino acid sequence of the vascular adhesion protein-1.	
XX		
KW	Human; vascular adhesion protein-1; VAP-1; endothelial cell;	
KM	Lymphocyte; inhibition; amine oxidase.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 43	
FT	Misc-difference 47	/note- "potential O-glycosylation site"
FT	Misc-difference 679	/note- "potential O-glycosylation site"
FT	Misc-difference 679	/note- "potential O-glycosylation site"
FT	Misc-difference 137	/note- "potential O-glycosylation site"
FT	Misc-difference 137	/note- "potential N-glycosylation site"
FT	Misc-difference 232	/note- "potential N-glycosylation site"
FT	Misc-difference 294	/note- "potential N-glycosylation site"
FT	Misc-difference 592	/note- "potential N-glycosylation site"
FT	Misc-difference 618	/note- "potential N-glycosylation site"
FT	Misc-difference 666	/note- "potential N-glycosylation site"

GST: NABV 693:691
Human L1CAM F4-5
Fumonisin esterase
A fusion of aminop
GlutathioneStransf
Murine GST-modf fu
CAP-N fusion prote
GST: NABV 690:691
Amino acid sequenc
Mouse L1CAM F80 fu
GST/GI transport r
GST/GI transport r
Siglec-BMS-L3a cyt
GST/GI transport r
GST/GI transport r
GST/GI transport r
GST/GI transport r
GST/GI transport r
Turkey prolactin/G
GST/GI transport r
Glutathione-S-tran
Chicken Ng-CAM F4
GST/GI transport r
Amino acid sequenc
Chicken Ng-CAM Fm3
GST-Msp I methylas
GST/GI transport r
GST/GI transport r
Glutathione-S-tran
Trypanosoma cruzi
Trypanosoma cruzi
GST/human p57 fus1
GST/GI transport r
GST-SEP fusion pro
pGEX-Sep45 plasmid
GST/GI transport r


```

FT FT Peptide /note- "potential N-glycosylation site"
FT FT Peptide 74..78
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 170..174
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 259..262
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 264..277
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 323..329
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 359..361
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 368..375
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 384..393
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 562..566
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 597..601
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 624..637
FT FT Peptide /note- "V8 peptide"
FT FT Peptide 1..25
FT FT Peptide /note- "N-terminal tryptic peptide"
FT FT Domain 5..27
FT FT /note- "transmembrane domain"
XX PN MO9853049-A1.
XX PD 26-NOV-1998.
XX PF 22-MAY-1998; 98MO-F100429.
XX PR 23-MAY-1997; 97US-0862433.
XX PA (BIOT-) BIOTIE THERAPIES LTD.
XX PI Bono P, Jalkanen S, Salmi M, Smith DJ;
XX DR WPI; 1999-131690/11.
XX DR N-PSDB; AAX28632.
XX PT Nucleic acid encoding a novel human endothelial cell adhesion
XX PT protein designated VAP-1 - having an adhesive function and an amine
XX PT oxidase function useful for manipulating VAP-1 mediated binding of
XX PT endothelial cells to lymphocytes
XX PS Claim 1; Fig 1; 66pp; English.
XX CC This is the nucleotide sequence encoding the human vascular adhesion
XX CC protein-1 (VAP-1) used in the method of the invention. The method
XX CC involves manipulating VAP-1 mediated binding of endothelial cells
XX CC to lymphocytes which comprises inhibiting the enzymatic activity of
XX CC amine oxidase in endothelial cells, and potentiating the enzymatic
XX CC activity of endothelial cells.
SQ Sequence 763 AA;

Query Match 74.18; Score 3967; DB 20; Length 763;
Best Local Similarity 99.98; Pred. No. 0;
Matches 735; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 263 OGGGGGSEPSQLPHCPSPSPAPMTHPGOSQLFADLSREELTAVMRFLTORLPGELVDA 322
DB 28 RGGGGGSEPSQLPHCPSPSPAPMTHPGOSQLFADLSREELTAVMRFLTORLPGELVDA 87
QY 323 QARPSDNCVFSEVLEQLPKAALAHLDGSPPARALAIYFFGROPQPNVSELVVGPLP 382
DB 88 QARPSDNCVFSEVLEQLPKAALAHLDGSPPARALAIYFFGROPQPNVSELVVGPLP 147
QY 383 HPSYKRDVTVERRHGGLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHCCFFYHRRGN 442

```

```

DB 148 HPSYKRDVTVERRHGGLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHCCFFYHRRGN 207
QY 443 LVTMTTAPRGLOSGDRATMFGLYYINISGAGFLAHYGLLELLVNHKALDPAKMTIQKVFYQ 502
DB 208 LVTMTTAPRGLOSGDRATMFGLYYINISGAGFLAHYGLLELLVNHKALDPAKMTIQKVFYQ 267
QY 503 GRYYDSIAOLEAQFEAGLVNVLIPDNGTGSWSLSKSPVPPGAPPLQFYPGQPRFSVOG 562
DB 268 GRYYDSIAOLEAQFEAGLVNVLIPDNGTGSWSLSKSPVPPGAPPLQFYPGQPRFSVOG 327
QY 563 SRVASSLMTFSSFLGAFSSGRIFDVPRQGERLYEELISLOALAIYGGNSPAAMTTRIVOG 622
DB 328 SRVASSLMTFSSFLGAFSSGRIFDVPRQGERLYEELISLOALAIYGGNSPAAMTTRIVOG 387
QY 623 GFGMGKTYTTLTRGVDCPYATYVDHFLLESQAPKTRIDAFCVFEQNOGLPLRRHSD 682
DB 388 GFGMGKTYTTLTRGVDCPYATYVDHFLLESQAPKTRIDAFCVFEQNOGLPLRRHSD 447
QY 683 YSHYFGGLAETVLVRSMSSTLNDYVMDTVFHPGSAIEIRFYATGYISSAFLGATGKY 742
DB 448 YSHYFGGLAETVLVRSMSSTLNDYVMDTVFHPGSAIEIRFYATGYISSAFLGATGKY 507
QY 743 GNOVSEHTLGTVTHSHAFVDDVAGLEMMVVAEDVFPMAVPMSPEHQLDRLOYTRK 802
DB 508 GNOVSEHTLGTVTHSHAFVDDVAGLEMMVVAEDVFPMAVPMSPEHQLDRLOYTRK 567
QY 803 LLEMEBOAFLVGSATPRYLYLIASNSHKNMGHPRGYRIOMLSFAGEPLPONSSMARGF 862
DB 568 LLEMEBOAFLVGSATPRYLYLIASNSHKNMGHPRGYRIOMLSFAGEPLPONSSMARGF 627
QY 863 ERYOLAVTORKEEPPSSSVFNONDMPAPTVDFSDFTINNETIAGKDLVAVMTAGFLHIP 922
DB 628 ERYOLAVTORKEEPPSSSVFNONDMPAPTVDFSDFTINNETIAGKDLVAVMTAGFLHIP 687
QY 923 AEDIPNTVTYVNGVGFLLRPINFFDEDPSTYSADSIYFRDQDAGAEVNPPLACTPQAAA 982
DB 688 AEDIPNTVTYVNGVGFLLRPINFFDEDPSTYSADSIYFRDQDAGAEVNPPLACTPQAAA 747
QY 983 CAPDLPAFSHGGSFHN 998
DB 748 CAPDLPAFSHGGSFHN 763

RESULT 2
AAU84261
ID AAU84261 standard; Protein; 729 AA.
XX AC AAU84261;
XX DT 08-MAY-2002 (first entry)
XX DE Human endometrial cancer related protein, AOC2.
XX KW Human; endometrial cancer; differential expression;
XX KW DNA microarray; protein microarray.
XX OS Homo sapiens.
XX PN MO200209573-A2.
XX PD 07-FEB-2002.
XX PF 31-JUL-2001; 2001MO-US24104.
XX PR 31-JUL-2000; 2000US-221735P.
XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
XX PI Mutter GL;
XX DR WPI; 2002-179967/23.
XX DR N-PSDB; ABK35481.

```

PT Diagnosing endometrial cancer comprises determining expression of
PT nucleic acid molecules or expression products that are differentially
PT expressed in normal and malignant endometrium -
XX
Claim 33; Page 131-134; 233pp; English.
PS
XX

The invention relates to diagnosing endometrial cancer in a subject suspected of having endometrial cancer comprising determining the expression of a set of nucleic acid molecules or expression products in an endometrial sample suspected of being cancerous, where the set of nucleic acid molecules comprises at least 2 nucleic acid molecules selected from 50 fully defined sequences as given in the specification. The nucleic acids are used as an array of at least 2 of the 50 nucleic acids bound to a solid substrate. Also included is a solid-phase protein microarray comprising at least 2 antibodies or its antigen binding fragments, that specifically bind at least 2 different polypeptides from the 50 fully defined sequences as given in the specification, fixed to a solid substrate. The methods and arrays are useful for the diagnosis of endometrial cancer, selecting and monitoring treatment regimes and identification of lead compounds useful for the treatment of endometrial cancer. The present sequence is one of 50 proteins differentially expressed between cancerous and non-cancerous samples.

Sequence 729 AA;

Query Match	46.3%	Score 2480.5	DB 23	Length 729
Best Local Similarity	64.1%	Pred. No. 4.1e-223		
Matches 468; Conservative	86;	Mismatches 145;	Indels 31;	Gaps 3;

Qy	268	GEPSQLHCSCSVSSAPRMTNPGSQJFADLSRELTAVNRFTLQRIQSGPLVDAACAPRS	327
Db	29	GGSSQPRHCSCSVSHRACPWRHPGSOJFADLSRELTAVNRFTLQRIQSGPLVDAACAPRS	88
Qy	328	DNCVSEVLOLPRKAALAHLDRCSPPRAREALIAVEFGKOPOPNSELVGLPRHESYM	387
Db	89	DNCJFSEVLOLPRKAALAHLDRCSPPRAREALIAVEFGKOPOPNSELVGLPRHESYM	148
Qy	388	RDVYVEHGGGLPVHRPRVLEQJELDDOMIFNELPQASGLDHHCFCYHHRGRNLVTMT	447
Db	149	RDVYVEHGGGLPVHRPRVLEQJELDDOMIFNELPQASGLDHHCFCYHHRGRNLVTMT	206
Qy	448	TAPRGLOSGDRATFGLYUNISGAFPLAHNVGJELLVNHRALDPARMTIOKVEFGKRYD	507
Db	207	ATPRGLSREMTTIGLYHNISGJGLPLHRYGJELLDHRAIDRPAHNTYOQVFLGHYXA	266
Qy	508	SLADLEKQFAGJLVNVTLPDNGTGGSSWSLSPVRPGRAPRLOFPOGPRFSSVOGSNVAS	567
Db	267	DLGLEEFKSGRLVEYRVLP.LPRPNGASLSRNSRSPGLRPLP.FSPGSGSYVOGJNVVS	326
Qy	568	SLATFSFGJLGFSPRJEDVRFOGERYLVUEISLOEALATIGNSPRAMTTRVYDGGFGMG	627
Db	327	SLMFSTGCHGFSGLRFLPDVRFOGERLAIEVSQDECSYIIGABSPKMLTRLYDSSFGJG	366
Qy	628	KYTPPLTRGYDCPLATVYDMHLLFLESQAPRXTTIDAFCEVQUNQJLRLRRHSDLYSHYF	687
Db	387	RNSGLVRCGYDCPRQATMYVDIHILVYGKAVQJLPRGAVCEFEAGJPLRLRHNYLQNHFY	446
Qy	688	GGJLEJLVNRSMTLLNYYVMQTVYHNRGSATIRRYALGJISSALFEGATG--KYGNQ	745
Db	447	GGJLSSALVNVASVSVCNENTIMQFVLYPNGALEGRHAGYINTAFLKGGEBGLTGNR	506
Qy	746	VSEHTLGTVHTSHAFKJDDVLAJLMMVAAEDHVFVPMVAPMSPBJOLRLOYTRKJLE	805
Db	507	VGEHVLGTVHHAHFKJDDVLAJLMMVAAEDHVFVPMVAPMSPBJOLRLOYTRKJYLG	566
Qy	806	MEQOAEFLVGSATERPITYLASNSHNSKWKHPRGYRIOMLSFAGEPLPONSSMARGFSWERY	865
Db	567	KEDJLAFSGJPLRPRYLYLASNOJNANMOHQG-----Y	598
Qy	866	QJAVTQKREEPSSSSVYQNDMPAPRYVDSDFINNETJAGKJLVAVNTYAGFLIPIAED	923
Db	600	QJAVTQKREEPSOSSSIYHONDITPVTAFJDNINNTJLGGEDJLVAVNTYSFLIPIAED	659

QY	986	DLPFESHGFF	995
QY	926	IPPTVYIGNNGVGEFLRYNFEDEDPSSY	SADSIYFRDQDQAGCEVPLGLTPOAAACAP 985
Db	660	IPNVTYLGNNGVGEFLRYNFEDEDP	SIFRSGSYVEFKGDQAGLCSINPACLPDLAACVP 719
QY	720	DLPFESHGFF	729

```
QY      986 DLPFSGGF 995
      ||| ||: ||
Db      720 DLPFSYHG 729
```

RESULT 3

ID AU10670 standard; Protein; 447 AA.

AC AAU10670;

DT 14-FEB-2002 (first entry)
 YY

Mouse LICAM Fn4-5 fusion polypeptide.

KM Neurite outgrowth; fibronectin Type III repeat; cell adhesion molecule
KM Fm4-5; neurone; peripheral nerve damage; trauma; gliosis;
KM Infarction; degenerative disease; malignant disease; antibacterial;
KM central nervous system lesion; virucide; antiapoptosis; neurotropic;
KM neuroprotective; mouse; LIGM; mutant; muten.

OS	Mus sp.
OS	Synthetic.

US6313265-B1

PD 06-NOV-2001.

PF 24-JUL-1995; 95US-0506296

PR 24-JUL-1995; 95US-0506296

PA (Scripps Res Inst.

PI Phillips G, Cunningham BA, Crossin KL;

DR' WPI; 2002-017011/02.

PT Polypeptide for promoting neurite out-growth useful for treating PT diseases such as inflammation, Parkinson's disease, trauma, comprises PT fibronectin type III repeats derived from a family of cell adhesion PT molecules -

PS Claim 3; Fig 25; 132pp; English.

The present invention relates to polypeptides that promote neurite growth. The polypeptides contain fibronectin type III repeats derived from a family of cell adhesion molecules (CAMs). The polypeptides of the invention include the F80, 3-5, and 4-5 regions of the CAM family members chicken Ng-CAM, chicken Nr-CAM, mouse L1CAM and human L1CAM. The polypeptides of the invention are useful for promoting neurite outgrowth of neuronal cells in vitro e.g. in a cell culture system, or in vivo for treating disorders such as peripheral nerve damage associated with physical or surgical trauma, infarction, bacterial or viral infections, toxin exposure, degenerative disease, malignant disease that affects peripheral or central neurones, or in surgical or transplantation methods in which new neuronal cells from brain, spinal cord or dorsal root ganglia are introduced and require stimulation of neurite outgrowth from the implant and innervation into the recipient tissue, where the diseases include central nervous systems lesions, gliosis, Parkinson's disease, Alzheimer's disease, gliotic response or inflammation. The present sequence represents mouse L1CAM Fnd-5 fusion polypeptide.

Sequence 447 AA;

Query Match	22.0%;	Score 1180;	DB 23;	Length 4477;
Best Local Similarity	69.7%;	Pred. No. 2e-101;		

Matches 251; Conservative 9; Mismatches 50; Indels 50; Gaps 10;

```

QY 38 SPIIGYWKIKGLVOTRLLLEYLEEKYEHLIERDEGKMKKKFELGLEPNNLPYYIDG 97
   |||||||
Db 2 SPIIGYWKIKGLVOTRLLLEYLEEKYEHLIERDEGKMKKKFELGLEPNNLPYYIDG 61
QY 98 DVKLTQSMATIRYIADKNNMLGSPKERAISMLEGAVLDIRGVSRAYSKDFFTLKVD 157
   |||||||
Db 62 DVKLTQSMATIRYIADKNNMLGSPKERAISMLEGAVLDIRGVSRAYSKDFFTLKVD 121
QY 158 FLTKLPEMLKMFEDRLSKITYLNGDVTHTPDMFLYDALDVVLYMDPMLDAFPKLVSRFK 217
   |||||||
Db 122 FLTKLPEMLKMFEDRLSKITYLNGDVTHTPDMFLYDALDVVLYMDPMLDAFPKLVSRFK 181
QY 218 RIEAIPQIDKYLKSKSYIAMPLOGQWATFFGGDHPKKSQSGEALFQGGDGEPSQLP-HC 276
   |||||||
Db 182 RIEAIPQIDKYLKSKSYIAMPLOGQWATFFGGDHPKKSQSGEALFQGGDGEPSQLP-HC 237
QY 277 PSVSPSAQPMW-----HP--GQS--QLFADLSREELTAVMRFLTORLGP 316
   |||||||
Db 238 QSDTSLHLHMQPRLSHNGVLTYGLLSYHPVEGSEKQGLFNLSDPELRT--HNLT-NLNP 294
QY 317 GLVDAQAARPSDNCVFSELDLPKKAALAHLDGSPPPA--REALIAVFFGROPQPNVS 374
   |||||||
Db 295 DL-----QYRFOLO-----AVTQOGGPGQAIIVREGGTWALFGKPDGNGIS 334

```

RESULT 4

AAU10662
ID AAU10662 standard; Protein: 644 AA.

AC AAU10662;

DT 14-FEB-2002 (first entry)

XX Chicken Ng-CAM F80 fusion polypeptide.

XX Neurite outgrowth; fibronectin type III repeat; cell adhesion molecule;

XX F80; neurone; peripheral nerve damage; trauma; gliosis;

XX Infarction; degenerative disease; malignant disease; antibacterial;

XX Central nervous system lesion; virucide; antiparkinsonian; nootropic;

XX Neuroprotective; antiinflammatory; chicken; Ng-CAM; mutant; mutein.

OS Gallus sp.

OS Synthetic.

PN US6313265-B1.

PD 06-NOV-2001.

PF 24-JUL-1995; 95US-0506296.

PR 24-JUL-1995; 95US-0506296.

PA (SCRI) SCRIPPS RES INST.

PI Phillips G, Cunningham BA, Crossin KL;

DR WPI; 2002-017011/02.

PT Polypeptide for promoting neurite out-growth useful for treating

PT diseases such as inflammation, Parkinson's disease, trauma, comprises

PT fibronectin type III repeats derived from a family of cell adhesion

PS molecules -

XX Example 2; Fig 17; 132pp; English.

XX The present invention relates to polypeptides that promote neurite

XX growth. The polypeptides contain fibronectin type III repeats derived

XX from a family of cell adhesion molecules (CAMs). The polypeptides of the

XX invention include the F80, 3-5, and 4-5 regions of the CAM family

XX members chicken Ng-CAM, chicken Nr-CAM, mouse L1CAM and human L1CAM.

XX The polypeptides of the invention are useful for promoting neurite

outgrowth of neuronal cells in vitro e.g. in a cell culture system, or
in vivo for treating disorders such as peripheral nerve damage
associated with physical or surgical trauma, infarction, bacterial or
viral infections, toxin exposure, degenerative disease, malignant
disease that affects peripheral or central neurones, or in surgical or
transplantation methods in which new neuronal cells from brain, spinal
cord or dorsal root ganglia are introduced and require stimulation of
neurite outgrowth from the implant and innervation into the recipient
tissue, where the diseases include central nervous systems lesions,
gliosis, Parkinson's disease, Alzheimer's disease, gliotic response or
inflammation. The present sequence represents chicken Ng-CAM F80 fusion
polypeptide.

SO Sequence 644 AA;

Query Match 21.9%; Score 1175.5; DB 23; Length 644;

Best Local Similarity 63.5%; Pred. No. 9.5e-101;

Matches 250; Conservative 19; Mismatches 52; Indels 73; Gaps 10;

```

QY 38 SPIIGYWKIKGLVOTRLLLEYLEEKYEHLIERDEGKMKKKFELGLEPNNLPYYIDG 97
   |||||||
Db 2 SPIIGYWKIKGLVOTRLLLEYLEEKYEHLIERDEGKMKKKFELGLEPNNLPYYIDG 61
QY 98 DVKLTQSMATIRYIADKNNMLGSPKERAISMLEGAVLDIRGVSRAYSKDFFTLKVD 157
   |||||||
Db 62 DVKLTQSMATIRYIADKNNMLGSPKERAISMLEGAVLDIRGVSRAYSKDFFTLKVD 121
QY 158 FLTKLPEMLKMFEDRLSKITYLNGDVTHTPDMFLYDALDVVLYMDPMLDAFPKLVSRFK 217
   |||||||
Db 122 FLTKLPEMLKMFEDRLSKITYLNGDVTHTPDMFLYDALDVVLYMDPMLDAFPKLVSRFK 181
QY 218 RIEAIPQIDKYLKSKSYIAMPLOGQWATFFGGDHPKKSQSGEALFQGGDGEPSQLP-HC 277
   |||||||
Db 182 RIEAIPQIDKYLKSKSYIAMPLOGQWATFFGGDHPKKSQSGEALFQGGDGEPSQLP-HC 238
QY 278 SVSP-----SAQPTHPGQSOLFADLSREELTAVMRFLTORLGP 317
   |||||||
Db 229 AEDPPFPFVALTYGGDARGALLGLRPM-----SRQL-RVLVFNQGGDP- 284
QY 318 LVDAQAARPSDNCVFSE--LQLPKKAALAHLDGSPPPAREALATY---FFGR----- 367
   |||||||
Db 285 -----PSEPIAFEPREGVGPPEELRVERLD-----DTLVSVERRTFRKSINGY 329
   |||||||
QY 368 -----QPQNVSELVVGPLPHPSYMRDVTVERHG 396
   |||||||
Db 330 VLRQGVVERG-----SALPGGSVLRDPQCDLRG 357

```

RESULT 5

AAAY21557
ID AAY21557 standard; Protein: 536 AA.

AC AAAY21557;

DT 03-AUG-1999 (first entry)

XX GS-GC6 fusion protein.

XX Cellular senescence; modulator; GC6 gene; senescent gene expression;

XX pGC6; human; fusion protein.

OS Homo sapiens.

PN WO9925878-A2.

PD 27-MAY-1999.

PF 19-NOV-1998; 98WO-US24996.

PR 19-NOV-1997; 97US-0974180.

PA (GERO-) GERON CORP.

PI Funk W:
 XX WPI: 1999-347496/29.
 DR N-PSDB: AAX77118.
 XX
 PT New human GC6 gene, useful for identifying agents for treating
 XX diseases and/or conditions associated with cell senescence
 PS
 XX Disclosure: Page 29-30; 79pp: English.

CC The invention relates to methods for modulating and identifying cellular
 CC senescence. Recombinant expression vectors comprising a recombinant
 CC polynucleotide corresponding to a polynucleotide in a human GC6 gene, are
 CC useful for altering senescent gene expression. The vectors and host cells
 CC comprising the vectors are useful for identifying agents that prevent or
 CC modulate senescent gene expression. The polynucleotides are useful for
 CC producing the protein, pGC6 and nucleic acid derivatives. The proteins
 CC encoded are useful for raising antibodies specific for pGC6, which are
 CC useful for isolating pGC6, and for detecting cells comprising pGC6 in
 CC complex cell mixtures. The characterization of the polynucleotides enable
 CC the identification of therapeutic agents that identify and distinguish
 CC between young and senescent cells. This enables treatment of aging
 CC diseases induced or exacerbated by cellular senescence.
 XX

Sequence 536 AA:

Query Match 21.9%; Score 1175; DB 20; Length 536;
 Best Local Similarity 49.6%; Pred. No. 7.9e-101;
 Matches 279; Conservative 41; Mismatches 123; Indels 120; Gaps 19;

QY 38 SPILGWTAKGLVOPRLLLEKEEYERDEGDKMRNKEFLGEPNLPYIDG 97
 DB 2 SPILGWTAKGLVOPRLLLEKEEYERDEGDKMRNKEFLGEPNLPYIDG 61
 QY 98 DVKLTOSMAIIRYIADKHNMLGSGPERAEISLMEGAVDIRGVSRIRYSKPEFLKVD 157
 DB 62 DVKLTOSMAIIRYIADKHNMLGSGPERAEISLMEGAVDIRGVSRIRYSKPEFLKVD 121
 QY 158 FLSTKLPEMLKMFEDRLSHNTYINGDHTVHPDFMLYDALDVLVYMDMCLDAPEFLKVSFK 217
 DB 122 FLSTKLPEMLKMFEDRLCHNTYINGDHTVHPDFMLYDALDVLVYMDMCLDAPEFLKVSFK 181
 QY 218 RLEAIPQIDKYLKSSKYTIAMPLOQWATFGGDPHPKSQSEALFOGGGGEPSQLPHCP 277
 DB 182 RLEAIPQIDKYLKSSKYTIAMPLOQWATFGGDPHPKSPD---LIEG--RGIPRNSARBD 235
 QY 278 SVSPSAQPTHPGQSGQLFADLSREELTAVMRFLTORLGGGLV----- 319
 DB 236 YFTNARELKKDAQODYHLEYVAMENSTHTTIEFTRELHCDINDKSIDTSTVAVIWAYHH 295
 QY 320 -DAQAARP-----SDNCVFSEVELQPPKAALAHLDGRSPPARALATVEFGHQPQPNYS 374
 DB 296 EDAGEGAPFYHDSNRSKTLRLNPEKTSVL-----STALYF----- 333
 QY 375 ELVGPRLPSTYMRDYTVRRHCGRLPYHRRPVLFOE---YLDIDOMIFNRELIPQASGLIH 431
 DB 334 DLVNDVPIPN--KDTTYCOMFKLP-----VFQKHNVIKVEPIYIGNH---ESLVH 381
 QY 432 HCCFYKHRGRNLVTMTTARGLSGSDRA-----TWPGLYYISAGNF- 473
 DB 382 HLLLYO-----CSNNFNDVLESIGECYHPNMPDAFLTCEYVIFAW-----ALGEGFS 430
 QY 474 FLHNHGLLELVNHNKALDPARMTIOKVYOGRYUDS-----LAOLEAOFEGALVYN 522
 DB 431 YPPHVGLSL---GTPLDP-HYVLELVHYDNPYTEEGLLDINSGLRLTYTMDIRKYDAGVIE 486
 QY 523 VVLIDNGTGGSM-SLKSVPVPG 544
 DB 487 -----AGLMVSLFHTIPPG 500

RESULT 6
 AAW47004

ID AAW47004 standard; Protein: 514 AA.
 XX
 AC AAW47004;
 XX
 DT 13-AUG-1998 (first entry)
 XX
 DE Glutathione-S-transferase and hTERT fusion protein 4.
 XX
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;
 XX prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.
 XX
 OS Synthetic.
 XX
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT Region 1..220 "glutathione-S-transferase fragment"
 FT Region 237..514
 FT Region /note- "hTERT protein fragment"
 PD GB2317891-A.
 XX
 PD 08-APR-1998.
 XX
 PF 01-OCT-1997; 97GB-0020890.
 XX
 PR 14-AUG-1997; 97US-0915503.
 PR 01-OCT-1996; 96US-0724643.
 PR 18-APR-1997; 97US-0844419.
 PR 25-APR-1997; 97US-0846017.
 PR 06-MAY-1997; 97US-0851843.
 PR 09-MAY-1997; 97US-0854050.
 PR 14-AUG-1997; 97US-0911312.
 PR 14-AUG-1997; 97US-0912951.
 XX
 PA (GERO-) GERON CORP.
 PA (UYTE-) UNIV. TECHNOLOGY CORP.
 XX
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J;
 PI Morin GB, Nakamura T, Harley CB;
 XX
 DR WPI: 1998-171633/16.
 XX
 PS pure and recombinant human Telomerase Reverse Transcriptase and its
 XX variants - are useful in the diagnosis, prognosis and treatment of
 XX cell proliferation conditions especially cancer and ageing
 XX
 PS Example 6; Page 226-227; 387pp: English.

The present sequence represents a fusion protein from an example of the present invention which describes human telomerase reverse transcriptase (hTERT). The present invention also describes the following methods: (A) determining whether a test compound is a modulator of hTERT, by detecting the change in hTERT recombinant protein or polynucleotide, on administration of the compound; (B) preparation of recombinant telomerase by contacting a protein preparation of hTERT with a telomerase RNA component; (C) detection of the hTERT RNA or protein in a sample by binding a relevant probe to the sample and detecting the complex formed or in the case of RNA detection, amplifying the product and correlating the presence of complex or amplification product with presence of hTERT in the sample; and (D) increasing the proliferation of a vertebrate cell by increasing hTERT expression; and (E) the use of an agent that causes an increase in cell vertebrate cell proliferation to create a medicament that inhibits ageing. A protein preparation of hTERT and the polynucleotide encoding hTERT can be used in the manufacture of medicaments for inhibiting the effect of ageing or cancer. Inhibitors of telomerase activity can be used to treat conditions that are associated with high telomerase activity. A protein preparation of hTERT can also be used in the new methods.

Sequence 514 AA:

```

Query Match      21.9%; Score 1171.5; DB 19; Length 514;
Best Local Similarity 75.4%; Pred. No. 1.6e-100;
Matches 236; Conservative 7; Mismatches 39; Indels 31; Gaps 5;

QY 38 SPILGYWKIKGLVOPTRLLLEYLEEKEYEHLIERDEGKMKRKKFELGEPNLPYIDG 97
   |||||
DB 2 SPILGYWKIKGLVOPTRLLLEYLEEKEYEHLIERDEGKMKRKKFELGEPNLPYIDG 61
   |||||

QY 98 DVKLQSMATIRYIADKHNMLGSPKERAETSMLEGAVLDIRYGSRIAYSKDEFTLKYD 157
   |||||
DB 62 DVKLQSMATIRYIADKHNMLGSPKERAETSMLEGAVLDIRYGSRIAYSKDEFTLKYD 121
   |||||

QY 158 FLSKLPEMLKMFEDRLSHKTYLNGDVTHPDMLYDALDVLVLYMDPMCLDAFPKLVSEFK 217
   |||||
DB 122 FLSKLPEMLKMFEDRLCHKTYLNGDVTHPDMLYDALDVLVLYMDPMCLDAFPKLVSEFK 181
   |||||

QY 218 RIEAIPOIDKYLKSSKYIAMPLOGQATFGGDDHPKPSQ-----SQE 259
   |||||
DB 182 RIEAIPOIDKYLKSSKYIAMPLOGQATFGGDDHPKPSDLYPRGSRASVGSVTKATISLE 241
   |||||

QY 260 ALFQGDGDEPS--QLPHC--PSVSPSAQPTWTHP-----GQSOLFADLSREELTAVMR 308
   |||||
DB 242 GALGTRHSHSVGRQHNAGPSPSRPRPMDTPCPRYAETKHFLYSSGDKQELRP--S 299
   |||||

QY 309 FLTRQLGPGLVDA 321
   |||||
DB 300 FLTSLRPSLTGA 312
   |||||

RESULT 7
AAE02008
ID AAE02008 standard; Protein: 1252 AA.
XX
AC AAE02008;
XX
DT 31-JUL-2001 (first entry)
XX
DE Yeast cystathionine beta-synthase (CBS).
XX
KW Yeast; enzymatic cycling assay; cystathionine beta-synthase; CBS;
   cystathionine beta-lyase; CBL; cyclical interconversion;
   homocysteine.
OS Saccharomycetes cerevisiae.
XX
XX Key Location/Qualifiers
   FT Region 1..238
   FT /label= "Glutathione_S-transferase"
   FT /note= "GST fused to the amino-terminal of CBS"
   FT Region 239..745
   FT /label= "CBS-enzyme"
   FT /note= "Yeast cystathionine beta-synthase"
   FT MISC-difference 459
   FT /note= "Encoded by GAG"
   FT MISC-difference 460
   FT /note= "Encoded by GGA"
   FT MISC-difference 531
   FT /note= "Encoded by ACC"
   FT Region 746..1252
   FT /note= "Repeated CBS region"
XX
PN W0200133187-A2.
XX
XX 10-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-US30518.
XX
XX 02-NOV-1999; 99US-0163126.
XX 10-MAY-2000; 2000US-0203349.
XX 01-NOV-2000; 2000US-0704036.
XX
XX (CATC-) CATCH INC.
XX

```

```

PI Kawasaki G, Webb HK, Owens J, Liedtke R, Forest D, Legaz M;
PI Lawson S;
XX
XX WPI: 2001-335847/35.
DR N-PSDB: AAD06034.
XX
PT Enzyme assays comprising using amounts of e.g. pyruvate and ammonia
PT produced during cyclical interconversion of homocysteine and
PT cystathionine to assess amounts of homocysteine and/or cystathionine in
PT a sample
XX
XX Claim 103; Page 75-79; 80pp; English.
PS
XX The present invention relates to an enzymatic cycling assays using
CC cystathionine beta-synthase (CBS) and cystathionine beta-lyase (CBL) to
CC assess the amount of homocysteine and/or cystathionine in a sample by
CC assaying the amount of pyruvate, ammonia, and/or nicotinamide adenine
CC dinucleotide (NAD+) produced during the cyclical interconversion of
CC homocysteine and cystathionine. The homocysteine reacts with L-serine to
CC form cystathionine in the presence of the enzyme CBS. The CBL catalyze
CC the conversion of cystathionine to homocysteine, pyruvate and ammonia.
CC The enzymatic cycling assay is less expensive and provides a higher
CC sample throughput than the diagnostic assays currently available.
CC The present sequence is yeast CBS having an amino terminal GST
CC fusion protein attached as a result of cloning into the bacterial
CC expression vector pGEX6P-2.
XX
SQ Sequence 1252 AA:

Query Match      21.8%; Score 1168; DB 22; Length 1252;
Best Local Similarity 76.2%; Pred. No. 1.4e-99;
Matches 234; Conservative 12; Mismatches 37; Indels 24; Gaps 4;

QY 38 SPILGYWKIKGLVOPTRLLLEYLEEKEYEHLIERDEGKMKRKKFELGEPNLPYIDG 97
   |||||
DB 2 SPILGYWKIKGLVOPTRLLLEYLEEKEYEHLIERDEGKMKRKKFELGEPNLPYIDG 61
   |||||

QY 98 DVKLQSMATIRYIADKHNMLGSPKERAETSMLEGAVLDIRYGSRIAYSKDEFTLKYD 157
   |||||
DB 62 DVKLQSMATIRYIADKHNMLGSPKERAETSMLEGAVLDIRYGSRIAYSKDEFTLKYD 121
   |||||

QY 158 FLSKLPEMLKMFEDRLSHKTYLNGDVTHPDMLYDALDVLVLYMDPMCLDAFPKLVSEFK 217
   |||||
DB 122 FLSKLPEMLKMFEDRLCHKTYLNGDVTHPDMLYDALDVLVLYMDPMCLDAFPKLVSEFK 181
   |||||

QY 218 RIEAIPOIDKYLKSSKYIAMPLOGQATFGGDDHPKPSQSQSEALPQG--GGGDEPSQ--- 272
   |||||
DB 182 RIEAIPOIDKYLKSSKYIAMPLOGQATFGGDDHPKPS--DEVLFGPLGSPGIGSTMT 240
   |||||

QY 273 -----LPHCPVSPSAQPTWTHPQSQQLPADLSREELTAVMRFLTORLPG 317
   |||||
DB 241 KSEQADSRHNVIDVGNTPPLIALKKLPKALGIRKQIYAKL---ELVNPGSINRIRAKS 297
   |||||

QY 318 LVDAQA 324
   |||||
DB 298 WVEEA 304
   |||||

RESULT 8
AAR79948
ID AAR79948 standard; Protein: 362 AA.
XX
XX AAR79948;
XX
XX 26-FEB-1996 (first entry)
XX
XX G17 antigen.
XX
XX Syphilis; G15 antigen; membrane antigen; glutathione-S-transferase;
XX GST; fusion protein; membrane antigen; diagnosis.
XX
XX Chimeric Treponema pallidum;
OS Chimeric unidentified bacterium.

```

```

XX Key Location/Qualifiers
FH Protein 1..228
FT /label= GST
FT Protein 229..362
FT /label= 17k_antigen
XX
XX EP670494-A2.
XX
XX PD 06-SEP-1995.
XX
XX PF 28-FEB-1995; 95SEP-0400423.
XX
XX PR 28-FEB-1994; 94JP-0054672.
XX
XX PA (FURE ) FUJI REBIO INC.
XX PA (FURE ) FUJI REBIO KK.
XX
XX PI Fujii N, Fujimura K, Okada M, Ueno E;
XX
XX DR WPI; 1995-303656/40.
XX
XX PT New assay for anti-Treponema pallidum antibody - useful for
XX PT diagnosis of syphilis.
XX
XX PS Disclosure; Page 10-11; 18pp; English.
XX
XX CC G17 antigen (AAR79948) is a fusion protein composed of glutathione-
XX CC S-transferase and the 17 kDa membrane antigen of Treponema
XX CC pallidum. It is obtd. by insertion of a DNA fragment coding for
XX CC the 17 kDa antigen into vector pWGA and expression in E. coli.
XX CC G17 and similarly produced G15 (AAR79947) are used for diagnosis
XX CC of syphilis.
XX
XX SQ Sequence 362 AA;

Query Match 21.8%; Score 1167.5; DB 16; Length 362;
Best Local Similarity 70.2%; Pred. No. 2.1e-100;
Matches 245; Conservative 15; Mismatches 56; Indels 33; Gaps 7;

QY 38 SPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKRNKKFELGEPNLPYYIDG 97
DB 2 SPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKRNKKFELGEPNLPYYIDG 61
QY 98 DVKLTOSMAIIRYIADKHNMLGSGPKERAEISMEGAVLDIRYGVSRIVASKDFETLKVD 157
DB 62 DVKLTOSMAIIRYIADKHNMLGSGPKERAEISMEGAVLDIRYGVSRIVASKDFETLKVD 121
QY 158 FLSKLPEMLKMFEDRLSHKTYTLNGDHYTHDPMLYDALDVLYLMDPCLDAFPLVSEFK 217
DB 122 FLSKLPEMLKMFEDRLCHKTYTLNGDHYTHDPMLYDALDVLYLMDPCLDAFPLVCFKK 181
QY 218 RIRAIPOIDKYLLSKSYIAMPLOGWQATFGGSDHPKSOQEAALFQGGDGEPSQLPHCP 277
DB 182 RIRAIPOIDKYLLSKSYIAMPLOGWQATFGGSDHPKSD---LVPRG-----SELVSC 222
QY 278 SVSPSAQPTWHPGOSQ-----LFADLSREELTAVARFETLORLGPGLVDAQAAPSDNC 330
DB 233 TVCP-----HAGKAKAEKVECALKGIFRGITLPAD-----CPG-IDITVTFFNADGT 278
QY 331 VRSVEYQLPPKAAALAHLDGSPPPAREA--LAIVFFGQPOPQNVSEL 376
DB 279 AQRVELALEKKSAPSPLYRTGTMVREDGIVELSLVSSQSKAPHEKEL 327

RESULT 9
AA96056
ID AA96056 standard; Protein; 658 AA.
AC AA96056;
XX
XX DT 05-DEC-2000 (first entry)

```

```

DE Human suppressor of fused, hSu(fu), GST fusion.
XX
XX KW Suppressor of fused; hSu(fu); PRO1280; human; hedgehog;
XX KW signal transduction; cancer; tumour suppressor; therapy; diagnosis;
XX KW glutathione S-transferase; GST.
XX
XX OS Chimeric - Homo sapiens.
XX OS Chimeric - Schistosoma sp.
XX
XX FH Key Location/Qualifiers
XX FT Protein 1..226
XX FT /label= GST
XX FT Protein 227..658
XX FT /label= hSu(fu)
XX
XX PN WO200052166-A1.
XX
XX PD 08-SEP-2000.
XX
XX PF 02-MAR-2000; 2000WO-US05746.
XX
XX PR 05-MAR-1999; 99US-0123090.
XX PR 25-MAY-1999; 99US-0135736.
XX
XX PA (GETH ) GENENTECH INC.
XX
XX PI De Sauvage FJ, Gurney AL, Murone M, Rosenthal A, Stone DM;
XX PI Wood WI.
XX
XX DR WPI; 2000-565502/52.
XX
XX PT New human suppressor of fused proteins and nucleic acids encoding them
XX PT useful e.g. as hybridization probes, in chromosome and gene mapping, in
XX PT the generation of antisense RNA and DNA
XX
XX PS Example 6; Fig 11; 67pp; English.
XX
XX CC The present sequence is that of a fusion between glutathione
XX CC S-transferase (GST) and human suppressor of fused (hSu(fu), see
XX CC also AA96054), a protein that functions as a negative regulator of
XX CC the hedgehog segment polarity pathway. hSu(fu) is highly expressed
XX CC in regions of active cell proliferation, and is probably a tumour
XX CC suppressor protein. hSu(fu) was expressed as a GST fusion protein
XX CC using vector pGEX in transformed bacterial cells, and used to study
XX CC the protein-protein interactions of hSu(fu). The invention provides
XX CC hSu(fu) nucleic acids (see AA50501), proteins, host cells, vectors,
XX CC antibodies, agonists and antagonists, as well as a method of
XX CC determining whether a disorder is modulated by hedgehog signalling.
XX
XX SQ Sequence 658 AA;

Query Match 21.8%; Score 1166; DB 21; Length 658;
Best Local Similarity 87.2%; Pred. No. 7.7e-100;
Matches 225; Conservative 6; Mismatches 23; Indels 4; Gaps 2;

QY 38 SPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKRNKKFELGEPNLPYYIDG 97
DB 2 SPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKRNKKFELGEPNLPYYIDG 61
QY 98 DVKLTOSMAIIRYIADKHNMLGSGPKERAEISMEGAVLDIRYGVSRIVASKDFETLKVD 157
DB 62 DVKLTOSMAIIRYIADKHNMLGSGPKERAEISMEGAVLDIRYGVSRIVASKDFETLKVD 121
QY 158 FLSKLPEMLKMFEDRLSHKTYTLNGDHYTHDPMLYDALDVLYLMDPCLDAFPLVSEFK 217
DB 122 FLSKLPEMLKMFEDRLCHKTYTLNGDHYTHDPMLYDALDVLYLMDPCLDAFPLVCFKK 181
QY 218 RIRAIPOIDKYLLSKSYIAMPLOGWQATFGGSDHPKSOQEAALFQGGDGEPSQLPHCP 277
DB 182 RIRAIPOIDKYLLSKSYIAMPLOGWQATFGGSDHPKSD---DVPRGSALRPSGAPG-P 237
QY 278 SVSPSAQPTWHPGOSQLF 295

```


CC According to the text of the specification, the C-terminal linker
 CC comprises 5 residues, however, only the sequence Aa Ser Ser END
 CC is given.
 CC The GST-derived protein domain allows rapid isolation using affinity
 CC chromatography to a solid phase contg. glutathione affixed to it.
 CC The Factor Xa specific cleavage site is provided when a subject
 CC fusion protein is produced in the vector pGEX-3X. The cleavage
 CC site is included to allow produ. of an NMBV structural protein
 CC free of the GST function domain.
 XX
 SQ Sequence 285 AA:
 Query Match 21.7%; Score 1164; DB 13; Length 285;
 Best Local Similarity 88.5%; Pred. No. 3e-100;
 Matches 224; Conservative 3; Mismatches 14; Indels 12; Gaps 3;
 QY 38 SPLIGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
 DB 2 SPLIGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
 QY 98 DVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRISAKDFETLKVD 157
 DB 62 DVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRISAKDFETLKVD 121
 QY 158 FLSKLPKMLKMFEDRLSHKTYLNGDHYTHPDMLYDALDVLYMDPCLDAPFKLYSFKK 217
 DB 122 FLSKLPKMLKMFEDRLCHKTYLNGDHYTHPDMLYDALDVLYMDPCLDAPFKLYCFKK 181
 QY 218 RIRAIPOIDKYLKSSKIYIAMPLOGMOATFGGDPHPKPSQSEALFQGGDGSEPSQLPHCP 277
 DB 182 RIRAIPOIDKYLKSSKIYIAMPLOGMOATFGGDPHPKPSD-----LIEG--RGIFN----- 229
 QY 278 SVSPSAQPTWHPG 290
 DB 230 SVSPSAQPTWHPG 242
 RESULT 12
 ID AAU10667 standard; Protein; 442 AA.
 AC AAU10667;
 DT 14-FEB-2002 (first entry)
 DE Human L1CAM Fn4-5 fusion polypeptide.
 XX
 KM Neurite outgrowth; fibronectin Type III repeat; cell adhesion molecule;
 KM Fn4-5; neurone; peripheral nerve damage; trauma; gliosis;
 KM Infectious; degenerative disease; malignant disease; antibacterial;
 KM central nervous system lesion; virucide; antiparkinsonian; neurotropic;
 KM neuroprotective; antiinflammatory; human; L1CAM; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 PN US6313265-B1.
 PD 06-NOV-2001.
 XX
 PF 24-JUL-1995; 95US-0506296.
 PR 24-JUL-1995; 95US-0506296.
 XX
 PA (SCRI) SCRIPPS RES INST.
 PI Phillips G, Cunningham BA, Crossin KL;
 DR WPI; 2002-017011/02.
 XX
 PT Polypeptide for promoting neurite out-growth useful for treating
 PT diseases such as inflammation, Parkinson's disease, trauma, comprises
 PT fibronectin type III repeats derived from a family of cell adhesion

PT molecules -
 XX
 PS Claim 3; Fig 22; 132pp; English.
 XX
 CC The present invention relates to polypeptides that promote neurite
 CC growth. The polypeptides contain fibronectin Type III repeats derived
 CC from a family of cell adhesion molecules (CAMs). The polypeptides of the
 CC invention include the F80, 3-5, and 4-5 regions of the CAM family
 CC members chicken Ng-CAM, chicken Nr-CAM, mouse L1CAM and human L1CAM.
 CC The polypeptides of the invention are useful for promoting neurite
 CC outgrowth of neuronal cells in vitro e.g. in a cell culture system, or
 CC in vivo for treating disorders such as peripheral nerve damage
 CC associated with physical or surgical trauma, infection, bacterial or
 CC viral infections, toxin exposure, degenerative disease, malignant
 CC disease that affects peripheral or central neurones, or in surgical or
 CC transplantation methods in which new neuronal cells from brain, spinal
 CC cord or dorsal root ganglia are introduced and require stimulation of
 CC neurite outgrowth from the implant and innervation into the recipient
 CC tissue, where the diseases include central nervous systems lesions,
 CC gliosis, Parkinson's disease, Alzheimer's disease, gliotic response or
 CC inflammation. The present sequence represents human L1CAM Fn4-5 fusion
 CC polypeptide.
 XX
 SQ Sequence 442 AA:
 Query Match 21.7%; Score 1164; DB 23; Length 442;
 Best Local Similarity 75.2%; Pred. No. 6.2e-100;
 Matches 236; Conservative 7; Mismatches 33; Indels 38; Gaps 6;
 QY 38 SPLIGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
 DB 2 SPLIGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
 QY 98 DVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRISAKDFETLKVD 157
 DB 62 DVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRISAKDFETLKVD 121
 QY 158 FLSKLPKMLKMFEDRLSHKTYLNGDHYTHPDMLYDALDVLYMDPCLDAPFKLYSFKK 217
 DB 122 FLSKLPKMLKMFEDRLCHKTYLNGDHYTHPDMLYDALDVLYMDPCLDAPFKLYCFKK 181
 QY 218 RIRAIPOIDKYLKSSKIYIAMPLOGMOATFGGDPHPKPSQSEALFQGGDGSEPSQLPHCP 276
 DB 182 RIRAIPOIDKYLKSSKIYIAMPLOGMOATFGGDPHPKPSD-----LVPRSGGHPALLEC 237
 QY 277 PSVSPSAQPT-----HP-----GOSQFADLSREE-----LRAVMR 308
 DB 238 QSNVTSLLRRQPPLSHNGVLTGYLSYHPDDEGKGSLFNLDPELRTINTLDSLHLR 297
 QY 309 F-----LTORLAPG 317
 DB 298 YRFQLQATTKEGPG 311
 RESULT 13
 ID AAY58917 standard; Protein; 1196 AA.
 AC AAY58917;
 DT 23-MAY-2000 (first entry)
 DE Fumonisin esterase-amino polyol amine oxidase fusion.
 XX
 KM Amino polyol amine oxidase; APAO; fumonisin esterase; mycotoxin;
 KM transgenic plant; detoxification; animal feed; silage;
 KM glutathione S transferase.
 XX
 OS Chimeric - Schistosoma japonicum.
 OS Chimeric - Bacterium.
 OS Chimeric - Exophiala splifera.
 XX
 PH key Location/Qualifiers


```

FT Peptide 1..229 /note="glutathione S transferase + polylinker"
FT Protein 230..1196 /note="fumisin esterase-K:trAPAO fusion"
FT Misc-difference 734 /note="extra lysine"
FT Peptide 721..733 /note="peptide linker"
XX WO200004159-A1.
XX
XX 27-JAN-2000.
XX
XX 08-JUL-1999; 99MO-US15454.
XX
XX 15-JUL-1998; 98US-0092936.
XX
XX 21-MAY-1999; 99US-0135391.
XX
XX (PION-) PIONEER HI-BRED INT. INC.
XX (CURA-) CURAGEN CORP.
XX
XX Duvick JP, Gilliam JT, Maddox JR;
XX
XX MPI; 2000-182425/16.
XX
XX N-PSDB; AA258406.
XX
XX New isolated polynucleotides, polypeptides useful for detecting and
XX degrading fumonisin or structurally related mycotoxin in processed
XX grain or in silage.
XX
XX Example 13; Page 129-131; 154pp; English.
XX
XX The present sequence is that of a fusion protein composed of
XX Schistosoma japonicum glutathione S transferase, and a bacterial
XX fumonisin esterase (BEST1) joined via a peptide linker to K:trAPAO
XX (see AA558909), a truncated, but functional, amino polyol amine
XX oxidase of E. spinifera. The construct was designed for expression
XX in Escherichia coli of a hybrid protein having BEST1 and APO
XX activity. The invention provides APO polynucleotides (see
XX AA258383-87) and polypeptides (see AA558900-05) of E. spinifera and
XX Rhinoclella atrovirens. The polynucleotides are used to
XX transform plant cells normally susceptible to Fusarium or other
XX toxin-producing fungus infection. Also provided are methods for
XX expressing APO, optionally as a fusion protein with fumonisin
XX esterase, in transgenic plants, prokaryotic and non-plant
XX eukaryotic systems.) Methods for detoxification of grain, grain
XX processing, silage, food crops and in animal feed and rumen
XX microorganisms are also disclosed.
XX
XX Sequence 1196 AA:
XX
XX Query Match 21.7%; Score 1163; DB 21; Length 1196;
XX Best Local Similarity 38.9%; Pred. No. 3.9e-99;
XX Matches 325; Conservative 55; Mismatches 188; Indels 268; Gaps 31;
XX
XX QY 38 SPILGYKIKGLVOPTRLLLEYLEEKEEHLVERDECDKMNKFEELGLEPPNLPYITDG 97
XX |||||||
XX DB 2 SPILGYKIKGLVOPTRLLLEYLEEKEEHLVERDECDKMNKFEELGLEPPNLPYITDG 61
XX
XX QY 98 DVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVIDIRGVSRIRAYSKDEFETLKV 157
XX |||||||
XX DB 62 DVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVIDIRGVSRIRAYSKDEFETLKV 121
XX
XX QY 158 FLKSLPMLKMFEDRLSHKRYTLNGDHTHPDMKYDALDVVLYNDPCLDAFPKLVSEFKK 217
XX |||||||
XX DB 122 FLKSLPMLKMFEDRLSHKRYTLNGDHTHPDMKYDALDVVLYNDPCLDAFPKLVSEFKK 181
XX
XX QY 218 RIBAIPOIDIKYLSKSKYIAMPLOGMATFGGDBPPKSSQSEALFQG----- 264
XX |||||||
XX DB 182 RIBAIPOIDIKYLSKSKYIAMPLOGMATFGGDBPPKSSQSEALFQG----- 238
XX
XX QY 265 -----GDGGEPSQLPHCPVSP-----SAQPTWHPGQSOLFADLSREELTAVMR 308
XX |||||||

```

```

DB 239 LGQVOGLAGDVMSFRGIPYAPPGVGLRMKPPQHARW-----AGVR 280
QY 309 FLVQRLGPGVLDAQAAP-----SDNCVFSVELQPPNA----- 342
DB 281 PATQ-FGSDCGAAYLKRKSLAPGVSEDCLY-LVMVAPSGAKPGQYVVMWVYGGFAGG 338
QY 343 -AALAHIDRGSPPARRELAIVFFGRQPPVNSLVVGPLPHEPYMDVYVERHGGLPY 401
DB 339 TAAMPYYD--GEALARQGVVVF-----NYRTNIIIGFFAHNG----- 374
QY 402 HRRPVLFOEYLDIDQMIENFRLPOASGLHHCCEYKRGRLVMTTAPRGLOS----- 455
DB 375 -----LSRESPTGTS-----GNGLLDILALAKRWQSNARAFG 407
QY 456 GD--RAIWFGLYINISGAGEFLHNVGDELLVNHKALDPARMTQKVFYQGRYYDS----- 508
DB 408 GDPGRVTVFGEBSAGASAIGLLTS-----PLSKGLFRAIILESPLTR 450
QY 509 -LAOLEAQFEAG-----LVNVVLIIDNGTGSWSLSKSPVPGPAPLPQFP 553
DB 451 PLATLADSAASGERLADLSRLRSTDPATLMARADARPARSRDLRRPPTGPIVDGHVLP 510
QY 554 QGPRFSVQGSRVASSLMTFSFGLGAFSGPRIFDVRFQGERLV-YEISLQ-----EALAIY 607
DB 511 QTDNAIATAAGOLAV--RVLIGTNADEG-RAFLGRAPMETPADYQAYLEAFOGQAAVA 567
QY 608 -----GNSPAAATTR-YVDGFGWG-----KY-----TRPLTR 635
DB 568 ACYPLDGRATPKEMVARIFGDQNFNRGVSFSEALVQGAPEVWRYQNGNTEGGRAPATH 627
QY 636 GVDCPYLATYVDMFLELSQAPKTRIRAPFCYFEQNO-GL-----PLRRHSDLYSH 685
DB 628 GAELPYV-----FGVFKDELDELDMPEEGPTPADRALGQLMSS 666
QY 686 YFGGLAETVLVVRMSLTLYNDYVMDTVFHPGSAIEIRFATGYISSAFLGATGK 741
DB 667 AWVAFK-----NDPRAGDALTPMA-----YSTG--KSTMTFPGPCR 701

```

RESULT 14
AA568853
ID AA568853 standard; Protein; 1196 AA.
XX
AC AA568853;
XX
DT 16-MAY-2000 (first entry)
XX
XX A fusion of aminopolylol amine oxidase/fumonisin esterase.
XX
KW Aminopolylol amine oxidase; APO; mycotoxin degradation; antimicrobial;
KW Plant fungal invasion; hydrogen peroxide; Fusarium; fungi;
XX fumonisin esterase.
XX
OS Synthetic.
OS Exophiala spinifera.
XX
FH Key Location/Qualifiers
FT Protein 1..229
FT Protein /note="GST and a linker"
FT Protein 230..721
FT Peptide /note="bacterial fumonisin esterase"
FT Peptide 722..733
FT Protein /note="spacer sequence"
FT Protein 7340..1196
FT /note="aminopolylol amine oxidase with an amino
terminal lys for optimized expression"

WO200004160-A1.
27-JAN-2000.
08-JUL-1999; 99MO-US15455.

PR 15-JUL-1998; 9805-0092936.
 XX (PION-) PIONEER HI-BRED INT INC.
 PA (CURA-) CURAGEN CORP.
 XX
 PI Crasta OR, Duvick J, Folkerts O, Gilliam JT, Maddox JR;
 DR WPI: 2000-182426/16.
 DR N-PSDB; AAZ60645.
 XX
 PT New nucleic acid encoding aminopolylol amine oxidase, used, e.g. to
 PT generate plants resistant to Fusarium -
 XX
 PS Example 13; Page 124-126; 145pp; English.
 XX
 CC The present sequence represents a fusion protein of GST/Exophiala
 CC splifera aminopolylol amine oxidase (APO)/fumonisin esterase.
 CC The APO enzyme has homology to the flavin containing amine oxidase
 CC family, that oxidise primary amine to an aldehyde or ketone, releasing
 CC ammonia and hydrogen peroxide. The APO enzyme degrades mycotoxins
 CC that promote fungal invasion of plants. Destruction of mycotoxins by
 CC APO generates, as a by-product, hydrogen peroxide which is itself an
 CC antimicrobial and stimulates the plants own defensive systems. The
 CC APO polynucleotides are used to generate plants (particularly maize)
 CC that are resistant to Fusarium or other fungi that produce mycotoxins
 CC and/or to degrade such mycotoxins (e.g. during ensiling); for
 CC recombinant production of APO polypeptides; as selection markers for
 CC plant transformation; and to isolate related sequences from other
 CC organisms. The APO polypeptides are used to degrade mycotoxins in
 CC plant materials, including expression in engineered bacteria and fungi,
 CC e.g. rumen microflora.
 CC
 XX
 SQ Sequence 1196 AA;
 Query Match 21.7%; Score 1163; DB 21; Length 1196;
 Best Local Similarity 38.9%; Pred. No. 3.9e-99;
 Matches 325; Conservative 55; Mismatches 188; Indels 266; Gaps 31;
 QY 38 SPILGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
 DB 2 SPILGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
 QY 98 DVKLTOSMAIIRYIADKHNMLGSPKERAISMLEGAVDIRGVSRIRYASKFETLKVD 157
 DB 62 DVKLTOSMAIIRYIADKHNMLGSPKERAISMLEGAVDIRGVSRIRYASKFETLKVD 121
 QY 158 FLTKLPEMLKMFEDRLSHKTYLNGDHYTHPDPEMLYDALVYLYMDMCLDAFPKLYSKK 217
 DB 122 FLTKLPEMLKMFEDRLCHHTYTLNGDHYTHPDPEMLYDALVYLYMDMCLDAFPKLYSKK 181
 QY 218 RIEAIPQIDIKYKSKSYTIMPLDQWQATFGGCDHPKSSQSEALFEG----- 264
 DB 182 RIEAIPQIDIKYKSKSYTIMPLDQWQATFGGCDHPKSSQSEALFEG----- 238
 QY 265 -----GDGEPQLPHCPSP-----SAQPMTHPGSQFLADLSRELTAYMR 308
 DB 239 LQVQGLADVMSFRGIPYAAAPRYVGLRKMPPOHARW-----AGVR 280
 QY 309 FLTORLPGVLDAQAARP-----SDNCVFSVELQLPKPA----- 342
 DB 281 PNTQ-FGSDCFGAAYLRKKSIAAGVSEDCLY-LNWAAPSAGKQGYRVWVWYGGGAGG 338
 QY 343 -AALAHLDNGSPPARALAIYFFGRQOPPNSELVVGPLPHPSYMRDVTVERHGCDLPY 401
 DB 339 TAMPYYD--GGAALARGVVYVF-----NYRTNIIIGFAHAG----- 374
 QY 402 HRRPVLFQEVLDIDQMIFFNELPQASGLHHCFFYKRGNTLYMTTAPRGLOS----- 455
 DB 375 -----LSRESPTGTS-----GNTGLDIIALALWVOSNARAFG 407
 QY 456 GD--RAATWFGLYNIGAGGFLLHVGLELLVNHKALDPAMTIOKVEYOGRYDS----- 508
 DB 408 GDGRVTYVGESAGASAGIIGLTS-----PLSKGLFRGAIIIESPOLTR 450

QY 509 -LAOLEAQFEAG-----LVNVVLIPDNGTGGSSLSKSPVPPGAPPIQERYP 553
 DB 451 PLATLADSAASGERLDADLSRLRSTDPATILMARADAPASRLRRPRTGIVDGHVLP 510
 QY 554 QGPRFSVGSKRVAASSLMTSFFGLGAFSPRITDVROGERLV-YEISLDQ-----EALAIY 607
 DB 511 QDSAAIAGOLAPV--RVLIGTNDEG--RAFLGRAPMETPADYQVYLEAQFGDQAAVA 567
 QY 608 -----GNSPAMTTR-YVDGFGMG-----KY-----TTPLTR 635
 DB 568 ACYPLDGRATPRKEMARIGDGNFNGVAFSEALVRQGAAPRWYOFNNTGGRAPATH 627
 QY 636 GVDCCPYLATYVDWHFLESQAPRTINDAFCEQNO-GL-----PLRRHSDLYSH 685
 DB 628 GAETIPV-----FGVEFLDELGLFDMPEGPPTPADRALGOLMS 666
 QY 686 YFGGLAETVLVVRSMSTLNDYVMDTVFHPGAIIRRYANGYISSAFLEGATGK 741
 DB 667 AMVRFKK-----NGDPAGDALTPA-----YSTG--KSTMTFEGGR 701
 RESULT 15
 ID AA017115 standard; Protein; 409 AA.
 XX AA017115;
 AC AA017115;
 XX
 DT 11-JUN-2002 (first entry)
 XX
 DE GlutathioneS-transferase/osteoclast differentiation factor fusion protein.
 XX
 KW Mouse; cathepsin K promoter; bone metabolic disease; osteoporosis;
 KW osteoclast differentiation factor; glutathione-S-transferase.
 XX
 OS Chimeric - Mus sp.
 XX
 PN JP2002034563-A.
 PD 05-FEB-2002.
 XX
 PF 19-JUL-2000; 2000JP-0218497.
 XX
 PR 19-JUL-2000; 2000JP-0218497.
 XX
 XX (SANY) SANKYO CO LTD.
 XX
 DR WPI: 2002-275721/32.
 XX
 PT Cathepsin K promoter useful for detecting a preventative agent for
 PT treating bone metabolic diseases e.g., osteoporosis -
 PS Disclosure; Page 14-15; 16pp; Japanese.
 CC
 CC The present invention provides the murine cathepsin K promoter. The
 CC sequence can be used in the identification of treatments and preventive
 CC agents for bone metabolic diseases such as osteoporosis. The present
 CC sequence is a fusion protein of the murine osteoclast differentiation
 CC factor and a glutathione-S-transferase described in the invention.
 CC
 XX
 SQ Sequence 409 AA;
 Query Match 21.7%; Score 1162.5; DB 23; Length 409;
 Best Local Similarity 87.6%; Pred. No. 7.5e-100;
 Matches 226; Conservative 1; Mismatches 24; Indels 7; Gaps 2;
 QY 38 SPILGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
 DB 2 SPILGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
 QY 98 DVKLTOSMAIIRYIADKHNMLGSPKERAISMLEGAVDIRGVSRIRYASKFETLKVD 157

Db 62 DVKLTQSMATIRYIADKHNMLOGGCPKERAEISMLEGAVIDIRYGVSRIRIAYSKDFETLKVD 121
QY 158 FLKLPMLKWFEDRLSHKTYLNGDHTHPDPMLYDALDVVLYMDPMCLDAFPKLVSEFK 217
Db 122 FLKLPMLKWFEDRLCHKTYLNGDHTHPDPMLYDALDVVLYMDPMCLDAFPKLVSEFK 181
QY 218 RIEAIPQIDKYLLKSSKYIAMPLOGMQATFGGSDHPPKSSQSEA-----LFGGSDG-GEP 270
Db 182 RIEAIPQIDKYLLKSSKYIAMPLOGMQATFGGSDHPPKSSQSEA-----LFGGSDG-GEP 241
QY 271 SQLPHCPSPVSPSAQPMTH 288
Db 242 SWLDVAQGRKPEAOPFAH 259

Search completed: May 20, 2003, 11:34:02
Job time : 90.2615 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 32.1643 Seconds

(without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357
Sequence: 1 MDWLRLFLMAAQSINAA.....QAAACAPDLPASHGFSHN 998

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents, AA.*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/6CTUS.COMB.pep.*
7: /cgn2_6/ptodata/2/1aa/6CTUS.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1180	22.0	447	US-08-506-296B-73	Sequence 73, Appl
2	1175.5	21.9	644	US-08-506-296B-65	Sequence 65, Appl
3	1175	21.9	536	US-08-974-186-20	Sequence 20, Appl
4	1171.5	21.9	514	US-08-974-349A-605	Sequence 605, App
5	1167.5	21.8	362	US-08-395-507-2	Sequence 2, Appl
6	1165	21.7	517	US-08-974-549A-606	Sequence 606, App
7	1164.5	21.7	443	US-08-506-296B-76	Sequence 76, Appl
8	1164	21.7	442	US-08-506-296B-70	Sequence 70, Appl
9	1163	21.7	1196	US-09-352-159-31	Sequence 31, Appl
10	1163	21.7	1196	US-09-352-168-31	Sequence 31, Appl
11	1162	21.7	272	US-08-910-820-4	Sequence 4, Appl
12	1162	21.7	397	US-08-327-874A-11	Sequence 11, Appl
13	1162	21.7	397	PCR-US94-09700-11	Sequence 11, Appl
14	1161	21.7	272	US-08-910-820-6	Sequence 6, Appl
15	1161	21.7	635	US-08-506-296B-71	Sequence 71, Appl
16	1158.5	21.6	426	US-08-737-248-4	Sequence 4, Appl
17	1158	21.6	439	US-08-506-296B-67	Sequence 67, Appl
18	1157.5	21.6	422	US-09-217-228-7	Sequence 7, Appl
19	1157.5	21.6	559	US-08-506-296B-56	Sequence 66, Appl
20	1157.5	21.6	647	US-08-305-764C-56	Sequence 56, Appl
21	1157	21.6	331	US-09-217-228-6	Sequence 6, Appl
22	1157	21.6	354	US-08-216-894-4	Sequence 4, Appl
23	1157	21.6	354	US-09-115-746-4	Sequence 4, Appl
24	1157	21.6	564	US-08-216-894-2	Sequence 2, Appl
25	1157	21.6	564	US-09-115-746-2	Sequence 2, Appl
26	1157	21.6	643	US-08-216-894-8	Sequence 8, Appl
27	1157	21.6	643	US-09-115-746-8	Sequence 8, Appl

28	1155.5	21.6	1140	US-08-471-112A-4	Sequence 4, Appl
29	1155	21.6	307	US-08-216-894-6	Sequence 6, Appl
30	1155	21.6	307	US-09-115-746-6	Sequence 6, Appl
31	1154.5	21.6	435	PCR-US95-04439-1	Sequence 1, Appl
32	1154	21.5	282	US-08-910-820-3	Sequence 3, Appl
33	1154	21.5	282	US-08-910-820-5	Sequence 5, Appl
34	1153	21.5	412	US-09-366-009-34	Sequence 34, Appl
35	1152.5	21.5	600	US-08-895-707-2	Sequence 2, Appl
36	1152.5	21.5	616	US-08-974-549A-602	Sequence 602, App
37	1152	21.5	632	US-08-506-296B-74	Sequence 74, Appl
38	1151.5	21.5	829	US-09-352-159-33	Sequence 33, Appl
39	1151.5	21.5	829	US-09-352-168-33	Sequence 33, Appl
40	1151	21.5	1205	US-09-352-159-29	Sequence 29, Appl
41	1151	21.5	1205	US-09-352-168-29	Sequence 29, Appl
42	1150.5	21.5	538	US-08-974-549A-602	Sequence 602, App
43	1150	21.5	472	US-08-216-894-10	Sequence 10, Appl
44	1150	21.5	472	US-09-115-746-10	Sequence 10, Appl
45	1149	21.4	542	US-08-506-296B-69	Sequence 69, Appl

ALIGNMENTS

```
RESULT 1
US-08-506-296B-73
; Sequence 73, Application US/08506296B
; Patent No. 6313265
; GENERAL INFORMATION:
; APPLICANT: Phillips, Greg
; APPLICANT: Cunningham, Bruce A.
; APPLICANT: Crossin, Kathryn L.
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF US
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESS: The Scripps Research Institute
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: U.S.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,296B
; FILING DATE: 24-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 488.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-506-296B-73
;
Query Match 22.0%; Score 1180; DB 4; Length 447;
Best Local Similarity 69.7%; Pred. No. 3.5e-107;
Matches 251; Conservative 9; Mismatches 50; Indels 50; Gaps 10;
OY 38 SPLIGWKIKGLVOPRLLLEYLEEKYEHLVERDGGDKWRNKKFELGLEFPPLPYIDG 97
DB 2 SPLIGWKIKGLVOPRLLLEYLEEKYEHLVERDGGDKWRNKKFELGLEFPPLPYIDG 61
```

Qy 38 SPILGWKIKGLVOPRRLLEYLEEKYEERHLYERDGGDKRRKKKDELGFENPLYIDG 97
 |||||
 Db 2 SPILGWKIKGLVOPRRLLEYLEEKYEERHLYERDGGDKRRKKKDELGFENPLYIDG 61

QY 38 SPILGYMKIGVQPTRLILEYLEEKYEHLEREDGKMKRKPELELEFPNLPYIDG 97
 |||||
 Db 2 SPILGYMKIGVQPTRLILEYLEEKYEHLEREDGKMKRKPELELEFPNLPYIDG 61

Tue May 20 13:18:58 2003

us-10-081-408-20.ra1

Page 3

```

QY 98 DVKTQSAAIIIRYADKNHNMGGSPKRAELSMIEGAVLDIRGVSTIASKPEFLTKD 157
Db 62 DVKTQSAAIIIRYADKNHNMGGSPKRAELSMIEGAVLDIRGVSTIASKPEFLTKD 121
QY 158 FLSLKPEMLKKFEDRLSHKTTYLNGDNVTHPDEFLYDALDVULLYMDPCLDAFPKLYSFK 217
Db 122 FLSLKPEMLKKFEDRLSHKTTYLNGDNVTHPDEFLYDALDVULLYMDPCLDAFPKLYSFK 181
QY 218 RIEAIPOLDIKLKSSTKIAMPLOGMAQTFGGDHPKKSQSQEAFLFGGDSGEPSQLPHCP 277
Db 182 RIEAIPOLDIKLKSSTKIAMPLOGMAQTFGGDHPKPSD----LIEG--RGIPRNSARED 235
QY 278 SVSPSAOPWTHPGOSQLFADLSREELTAVMKFGLTORLGRGLV----- 319
Db 236 YFTANNELKKAQODNHELEYAMENSHTIIIEFTRELHTCDIDKSTITDSTVAVIMAYNH 295
QY 320 -DAAQARP-----SDNCYFSVELDRPKAALANLDRSPPRAREALAYVEFGQRPQPNVS 374
Db 296 EDAGEAGKTYKDSNRKGSLSLNLPEKTSVL-----STALPRF----- 333
QY 375 ELVVGPLRHPBSYMDVTVERNHGRLPYRHRBPVLFQE---YLIDIDOMIFNELPQASGLTH 431
Db 334 DLVAGQDVPIRN--KDTTYWQOMFKIP-----VFQEKHNHVIKERVYIQRGH----ESLVN 381
QY 432 HCCSYKKNRGRLVMTTAPRGLSGGDA-----TWFGLYINISGAGF- 473
Db 382 HILLYQ-----CENNENDSVLESNGHCYHRNMDAFLTCTGYIFAW----AIGEGFS 430
QY 474 FLHVGLELIVNHNALPARPTIQKVEYQGRYDS-----LAQLEAQFEAGLVN 522
Db 431 YRPVHGSL---GPRDPR-HVYLLEVAYNDPYTEEGILINDSGRLFTYMOIRKRYDAVIE 486
QY 523 VLLPDPNGTGSW-SLKSPPVPPG 544
Db 487 -----AGLWVSLFHTTTPG 500

```

RESULT 4
 US-08-974-549A-605
 Sequence 605, Application US/08974549A
 Patent No. 6166178
 GENERAL INFORMATION:
 APPLICANT: Cech, Thomas R.
 APPLICANT: Lindner, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morin, Gregg B.
 APPLICANT: Hartley, Calvin H.
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: Human Telomerase Catalytic Subunit
 NUMBER OF SEQUENCES: 727
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/974,549A
 FILING DATE: 19-NOV-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997

```

PRIORITY APPLICATION DATA: US 08/846,017
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-00261005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 605:
SEQUENCE CHARACTERISTICS:
LENGTH: 514 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..515
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
and hprt protein fragment"
US-08-974-549A-605
/note="fusion protein composed of
glutathione-S-transferase sequence,
thrombin cleavage sequence, recognition
sequence for heart muscle protein
kinase residues introduced by cloning

```

	Query Match	Best Local Similarity	Matches	236; Conservative	21.9%; Score 1171.5; DB 4; Length 514; Pred. No. 3e-106; Mismatches 93; Indels 31; Gaps 5;
QY	38	SPILGYWKIKGLVOPTRLILLEYLEEKYEEDHLYERDEGDKWRNKKFELGLEFPNLPYYIDG	97		
Db	2	SPILGYWKIKGLVOPTRLILLEYLEEKYEEDHLYERDEGDKWRNKKFELGLEFPNLPYYIDG	61		
QY	98	DVKLTQSAIIRYIADKNHNLGGSPKEKAEIEMLGAVLDIYGYSRIRYSQDFETLKYD	157		
Db	62	DVKLTQSAIIRYIADKNHNLGGSPKEKAEIEMLGAVLDIYGYSRIRYSQDFETLKYD	121		
QY	158	FLSKLPELKKMFEDRLSKITYLNGDHTVHPDFMLYDALDYVLKMDPCLDAEPKLVSRFK	217		
Db	122	FLSKLPELKKMFEDRLSKITYLNGDHTVHPDFMLYDALDYVLKMDPCLDAEPKLVSRFK	181		
QY	218	RIEAIPQIDKYLKSSKSYLAMPLOGQATFFGGGDHPPKSO-----	259		-S0E
Db	182	RIEAIPQIDKYLKSSKSYLAMPLOGQATFFGGGDHPPKSDLYVRGSRASVGSVTKATSLD	241		
QY	260	ALFPGGDDGGES--QLPFC--PSVSPSKQPPWTHP-----	308		-GQSDLPADLSNEELTAAYR
Db	242	GALSGTRHSHSPVGRQHINAGPPSTSRPRPMDTPCPYVAETKHKHFLYSGGDEQLRP--S	299		

QY 309 FLNORLPGGLVDA 321
DB 300 FLNLSRLPSLTGA 312

RESULT 5

US-08-395-507-2
Sequence 2, Application US/08395507
Patent No. 5578456
GENERAL INFORMATION:
APPLICANT: Fujimura, Katsuya
APPLICANT: Ueno, Eiichi
APPLICANT: Fujii, No. 5578456yuk1
APPLICANT: Okada, Masahisa
TITLE OF INVENTION: Anti-treponema Pallidum Antibody
TITLE OF INVENTION: Immunoassay
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pollock, Vande Sande & Priddy
STREET: 1990 M Street, NW, Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,507
FILING DATE: 27-FEB-1995
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP Hei-6-54672
FILING DATE: 28-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Amerinick, Burton A.
REGISTRATION NUMBER: 24,852
REFERENCE/DOCKET NUMBER: 1581/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 331-7111
TELEFAX: (202) 293-6229
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 362 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: Protein
LOCATION: 1..362
OTHER INFORMATION: /note= "G17 Antigen"
US-08-395-507-2

Query Match 21.8%; Score 1167.5; DB 1; Length 362;
Best Local Similarity 70.2%; Pred. No. 4.1e-106;
Matches 245; Conservative 15; Mismatches 56; Indels 33; Gaps 7;

QY 38 SPILGYWKINGLVOPTRLLLEYLEEKEYEHLRYERDEGDKMNNKFELGLEPPNLPYYIDG 97
DB 2 SPILGYWKINGLVOPTRLLLEYLEEKEYEHLRYERDEGDKMNNKFELGLEPPNLPYYIDG 61
QY 98 DVKLTQSMATIRYADKHNMLGSSPKERAETSMLEGAVALDIRYGVSRVAYSKDEFTLKV 157
DB 62 DVKLTQSMATIRYADKHNMLGSSPKERAETSMLEGAVALDIRYGVSRVAYSKDEFTLKV 121
QY 158 FLNLSPEMLKMEEDRLSKKTYLNGDHVTHRPDMLYDAIDVYLYMDPMLDAPPKYVSRK 217
DB 122 FLNLSPEMLKMEEDRLSKKTYLNGDHVTHRPDMLYDAIDVYLYMDPMLDAPPKYVSRK 181

QY 218 RIEALPIDKYLKSKXIAMPLOGMOATFGGDHPKPSQSQELAFQGDGDEPSQLPHP 277
DB 182 RIEALPIDKYLKSKXIAMPLOGMOATFGGDHPKPSQ-----LVPRG-----SELYSCT 232
QY 278 SVSPSAQPMTHPGOSQ-----LFADLSREELTAVMRFLTORLPGGLVDAQARPSDNC 330
DB 233 TVCP-----HAKKAKAEKVECALKGIFRGTLPAAD-----CPG-IDTIVTFNADGT 278
QY 331 VFSEYLOLPPKAAALNLDGSPPPAREA---LAIVFGHPOPPNSEL 376
DB 279 AQKVELALEKKSAPSPITRYGTWVRREDGIVELSLVSSQSKAPHEKEL 327

RESULT 6

US-08-974-549A-606
Sequence 606, Application US/08974549A
Patent No. 6166178
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Hartley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 536
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885

FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 606:
SEQUENCE CHARACTERISTICS:
LENGTH: 517 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..517
OTHER INFORMATION: /note="fusion protein composed of
OTHER INFORMATION: glutathione-S-transferase sequence,
OTHER INFORMATION: thrombin cleavage sequence, recognition
OTHER INFORMATION: sequence for heart muscle protein
OTHER INFORMATION: kinase, residues introduced by cloning
OTHER INFORMATION: and hprt protein fragment
US-08-974-549A-606

Query Match 21.7%; Score 1165; DB 4; Length 517;
Best Local Similarity 66.9%; Pred. No. 1.3e-105;
Matches 239; Conservative 14; Mismatches 48; Indels 56; Gaps 6;

QY 38 SPILGWIKIGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYIDG 97
DB 2 SPILGWIKIGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYIDG 61
QY 98 DVKLTOSMAIIRYIADKHNMLGSGPKERAISMLEGAVIDIRGVSRIRVSKDFELKYD 157
DB 62 DVKLTOSMAIIRYIADKHNMLGSGPKERAISMLEGAVIDIRGVSRIRVSKDFELKYD 121
QY 158 FLSKLEPMLEKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPMCLDAFPLKVSFKK 217
DB 122 FLSKLEPMLEKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPMCLDAFPLKVSFKK 181
QY 218 RIFAIPOIDIKYLLSKSYTIAMPLOGWQATFGGSDHPPKSGSQEALFQGGDGEPSQLPHC 277
DB 182 RIFAIPOIDIKYLLSKSYTIAMPLOGWQATFGGSDHPPKSGSQEALFQGGDGEPSQLPHC 241
QY 278 -----SVSPSAQW-----PM-THPGQ 291
DB 242 RCRAVNSLSHREVLPATFVARLPGQGRVLQBGDPAAFRALVACLCVCPWMDARPPA 301
QY 292 SOLFADLS--REELTAVMRELTOR-----LQPGLVDAQAARPSDNCVSVELQLP 339
DB 302 APEFROVSCLEKELVAVLQRLCERGAKNVLAFFALDLGARGPPE-ATTSVASTYLP 357

RESULT 7
US-08-506-296B-76
Sequence 76, Application US/08506296B
Patent No. 6313265
GENERAL INFORMATION:
APPLICANT: Phillips, Greg
APPLICANT: Cunningham, Bruce A.
APPLICANT: Crossin, Kathryn L.
TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
TITLE OF INVENTION: CONTAINING FIBROBLAST TYPE III REPEATS AND METHODS OF USE
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute
STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
CITY: La Jolla
STATE: California
COUNTRY: U.S.
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506,296B
FILING DATE: 24-JUL-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 443 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
US-08-506-296B-76

Query Match 21.7%; Score 1164.5; DB 4; Length 443;
Best Local Similarity 74.4%; Pred. No. 1.1e-105;
Matches 236; Conservative 11; Mismatches 31; Indels 39; Gaps 7;

QY 38 SPILGWIKIGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYIDG 97
DB 2 SPILGWIKIGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYIDG 61
QY 98 DVKLTOSMAIIRYIADKHNMLGSGPKERAISMLEGAVIDIRGVSRIRVSKDFELKYD 157
DB 62 DVKLTOSMAIIRYIADKHNMLGSGPKERAISMLEGAVIDIRGVSRIRVSKDFELKYD 121
QY 158 FLSKLEPMLEKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPMCLDAFPLKVSFKK 217
DB 122 FLSKLEPMLEKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPMCLDAFPLKVSFKK 181
QY 218 RIFAIPOIDIKYLLSKSYTIAMPLOGWQATFGGSDHPPKSGSQEALFQGGDGEPSQLPHC 276
DB 182 RIFAIPOIDIKYLLSKSYTIAMPLOGWQATFGGSDHPPKSGSQEALFQGGDGEPSQLPHC 237
QY 277 PSVSPSAQW-----THPGQSLFADLSREELTA-VMRELT-----TORLQPGLVDAQAARPSD 328
DB 238 PTLDLSLTEMGSPTR-----NGVLTSTYILAFQPIINTHELCP----- 275
QY 329 NCYVSVELQLPKRAAL 345
DB 276 ---LVEIRIPANESSL 288

RESULT 8
US-08-506-296B-70
Sequence 70, Application US/08506296B
Patent No. 6313265
GENERAL INFORMATION:
APPLICANT: Phillips, Greg
APPLICANT: Cunningham, Bruce A.
APPLICANT: Crossin, Kathryn L.
TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
TITLE OF INVENTION: CONTAINING FIBROBLAST TYPE III REPEATS AND METHODS OF USE
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute
STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
CITY: La Jolla
STATE: California
COUNTRY: U.S.
ZIP: 92037
COMPUTER READABLE FORM:


```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506,296B
FILING DATE: 24-JUL-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 442 amino acids
TYPE: amino acid
      :
      : topology: linear
      :
      : MOLECULE TYPE: protein
      :
      : FRAGMENT TYPE: internal
      :
      : US-08-506-296B-70

```

Query Match	21.7%;	Score 1164;	DB 4;	Length 442;
Best Local Similarity	75.2%;	Pred. No. 1.3e-105;		
Matches 236; Conservative	7;	Mismatches 33;	Indels 38;	Gaps 6

```

Oy 38 SPIIGYMKIKLOVPTLYLLEYLEEKXEENLYERDEDEDKRNKKFTELGEFMYPIYIDG 97
Db 2 SPIIGYMKIKLOVPTLYLLEYLEEKXEENLYERDEDEDKRNKKFTELGEFMYPIYIDG 61
Oy 98 DVKLTQSMATITRYADKNHMLGSGPKERAETISMLEGAVLDIRYGVSTIAYSKOFETLYVD 157
Db 62 DVKLTQSMATITRYADKNHMLGSGPKERAETISMLEGAVLDIRYGVSTIAYSKOFETLYVD 121
Oy 158 FLSKLPEMLKMFEDRLSHKTYTLNGDHTHPDFMYLDALDVLYLMDPSCIDAPKLYSEKK 217
Db 122 FLSKLPEMLKMFEDRLCHKTYTLNGDHTHPDFMYLDALDVLYLMDPSCIDAPKLYSEKK 181
Oy 218 RLEAIPOIDKLYASKSKYIAMPLOSOWATFEGGDHPKYSQSOEALFQGGDGERSQLP-HC 276
Db 182 RLEAIPOIDKLYASKSKYIAMPLOSOWATFEGGDHPKPSD---LVPRGSGHBEALHLEC 237
Oy 277 PSVSPSAQPMW-----HP-----GQSOLFADLSHEE-----LTAVMR 308
Db 238 QSNMSTLLRMQRPILSHNGVLGYLVSYNHLRDEGGKQGLSFLRLRDLRTYHNLDLSPHLR 297
Oy 309 F-----LTQRLGPG 317
Db 298 YRFQLOAQATTKGPG 311

```

```

: RESULT 9
: US-09-352-159-31
: Sequence 31, Application US/09352159A
: Patent No. 6211434
: GENERAL INFORMATION:
: APPLICANT: Duwick, Jonathan P.
: APPLICANT: Gilliam, Jacob T.
: APPLICANT: Maddox, Joyce R.
: TITLE OF INVENTION: Amino Polyol Amine Oxidase
: TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
: FILE REFERENCE: 1134
: CURRENT APPLICATION NUMBER: US/09/352,159A
: CURRENT FILING DATE: 1999-07-12
: EARLIER APPLICATION NUMBER: 60/092,936
: EARLIER FILING DATE: 1998-07-25
: EARLIER APPLICATION NUMBER: 60/135,391
: EARLIER FILING DATE: 1999-05-21
: NUMBER OF SEQ ID NOS: 46
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO: 31

```

```

; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Unknown
US-09-352-159-31

```

Query Match	21.7%;	Score 1163;	DB 4;	Length 1196;
Best Local Similarity	38.9%;	Pred. No. 9.1e-105;		
Matches 325;	Conservative 55;	Mismatches 188;	Indels 268;	Gaps 31

QY	38	SPGVMKIKLVOPTRLLEYLEEKEENHLEYNEDDEDKRNKKFELGFEFNPYYIDG	97
Db	2	SPIGMKIKLVOPTRLLEYLEEKEENHLEYNEDDEDKRNKKFELGFEFNPYYIDG	61
QY	98	DVKLTQSMATIRYIADKRNMLGSGSPKERAISIMLEGAVLDIRGVSNIAYSKDETLKYD	157
Db	62	DVKLTQSMATIRYIADKRNMLGSGSPKERAISIMLEGAVLDIRGVSNIAYSKDETLKYD	121
QY	158	FLSKLPEMLKMFEDRLSKITTYLNDHNTYHDFMLYDLDVLYMDPMLCDAFKLYSFKK	217
Db	122	FLSKLPEMLKMFEDRLCHKTYLNDHNTYHDFMFLYDLDVLYMDPMLCDAFKLYSFKK	181
QY	218	RIEALPOLDKLKSSSKYIAMPLOGMOATFFGGDHPKPSQOEAFLFQG	264
Db	182	RIEALPOLDKLKSSSKYIAMPLOGMOATFFGGDHPKPSQOEAFLFQG	238
QY	265	-----GDGEPSSQLPHCBSPS-----SAOPWTHPGOSQLFADLSREBLIAYMR	308
Db	239	LGVOYGLAGDVMSRGRIPYAPAPVGLGKMRKPPHARPM-----AGVR	280
QY	309	FLTORLHSGGLVDAQAR-----SNOCVPSVBLQJPRKA-----	342
Db	281	PATQ-FGSDCFGAAYLRKGSILAPVSEBCLY-LVWMAPSGKKPGQYVVMWYGGGFAAG	338
QY	343	-AALAHDRGSPPREALATIVFFGROPORPVSELYVPLPHPSYMRDVTVERHGCLPY	401
Db	339	TAAMPYYD-GEALARGOVVYTF-----NYRNIIIGFAHPS-----	374
QY	402	HRRPVLFQOETDIDOMFNLRLPQASGLIHHCCTCYKHKRNLVYMTTAPRGLQS-----	455
Db	375	-----LSRSPSTGTS-----GNTGLDILALRNVOSNARAFG	407
QY	456	GD-RATWFGLYUNISGAGFLHHVUGLELVNHNKALDPARMTIOKVYOGGRYDS-----	508
Db	408	GDPCGVNTYFFGESAGASATGLLTS-----PLSKGLFRGAILSPBLTR	450
QY	509	-LAOLEAFEG-----LVNVVLLPDNGTGGSWLSKSPVPPGAPBLQFYP	553
Db	451	PLATLADSAASGERLADLRLRSTDPATLMAARADAPASRDRLRRPRTPGPIVDGHLP	510
QY	554	QGPRFSVSGSKYASSLMTFFSGCLAFSGPRFLDVRFOGERLY-YEISIQ-----EALAIY	607
Db	511	QDTSMAIAGOLAPY-RVYLGITNADBC-RAFLGRAPMETPADYQAVLEAOFDQAAVA	567
QY	608	-----GNSPAPMTTR-YVDGGFGMG-----XY-----TTPLTR	635
Db	568	ACYPLDGRATPREMARIRFGDNQFRGVSASFALVNGAPAVMYRQFNGNTEGGRATH	622
QY	636	GVDCCYLATYYDMHFLLESQAPKTIIRDAFCVFEONQ-GL-----PLRRHHSILYH	685
Db	628	GAEIFYV-----FGVFKLDELGLFMPWPEGCTPADRALGQLMSS	666
QY	666	YFGGLAEIVLVYRMSSTLLNTDYWDVYFHHSGAIEIRFYATYGISSAFJLGATGK	741
Db	667	AAVBRK-----NCDPADGALITWPA-----ISTG-KSTMFTGPEGR	701

RESULT: 10
US-09-352-168-31
; Sequence 31, Application US/09352168A
; Patent No. 6211935
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duvick, Jonathan P.

; APPLICANT: Folkerts, Otto
 ; APPLICANT: Gilliam, Jacob T.
 ; APPLICANT: Maddox, Joyce R.
 ; TITLE OF INVENTION: Amino Polyl Amino Oxidase
 ; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
 ; FILE REFERENCE: 08/75
 ; CURRENT APPLICATION NUMBER: US/09/352,168A
 ; CURRENT FILING DATE: 1999-07-12
 ; EARLIER APPLICATION NUMBER: 60/092,936
 ; EARLIER FILING DATE: 1998-07-25
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 31
 ; LENGTH: 1196
 ; TYPE: PRT
 ; ORGANISM: Unknown
 ; US-09-352-168-31

Query Match 21.7%; Score 1163; DB 4; Length 1196;
 Best Local Similarity 38.9%; Pred. No. 9.1e-105;
 Matches 325; Conservative 55; Mismatches 188; Indels 268; Gaps 31;

QY 38 SPILGWKIKGLVOPTRILLLEYLEKEEYERDEGDKRNKKFELGLEFPNLPYYIDG 97
 DB 2 SPILGWKIKGLVOPTRILLLEYLEKEEYERDEGDKRNKKFELGLEFPNLPYYIDG 61
 QY 98 DVKLTQSMATIRYIADKNNMLGSGPKERAEISMLEGAVLDIRYGSRIAYSKDEFTLKYD 157
 DB 62 DVKLTQSMATIRYIADKNNMLGSGPKERAEISMLEGAVLDIRYGSRIAYSKDEFTLKYD 121
 QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDPMYDALDVYLYNDPCLDAFPKLVSEFK 217
 DB 122 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDPMYDALDVYLYNDPCLDAFPKLVSEFK 181
 QY 218 RIEAIPOIDKYLKSSKIYAMPLOGMOATFGGSDHPKSSQSEALFQGG----- 264
 DB 182 RIEAIPOIDKYLKSSKIYAMPLOGMOATFGGSDHPKSSQSEALFQGG----- 238
 QY 265 -----GDGGEPSQLPHCPSPVSP-----SAQPMTHGQSOLFADLSREELTAVMR 308
 DB 239 LGQGVLAGDVMSFRGIPYAAPVYGLRWKRPQIARPM-----AGVR 280
 QY 309 FLTORLCPGLVDAQAARP-----SDNCVFSVELQLPKRA----- 342
 DB 281 PATD-FGSDCFGAAYLRKKSGLAPGVSEDCILY-LNVMA-PSGAKPCQYPMVWVYGGFAGG 338
 QY 343 -AALAHLDGRSPPARREALIVFGRQPPRVSELVVGPLPHSPYMDVTERHGGPLPY 401
 DB 339 TAAAPYYD--GEALARGVAVYTF-----NYRTNIIIGFFAHPC----- 374
 QY 402 HRRPVLFOEYLDIDQMTFNRFLPOASGLLHHCFCYKHRGNMVTMTTAPRGLQS----- 455
 DB 375 -----LSRESPTGTS-----GNYGLDILIALRVVQSNARAFG 407
 QY 456 GD--RAIWFGLIYVINGAGFELAHVGLLELVNKAALDPAWITQKVYQGRYDS----- 508
 DB 408 GDRPRVYVFGESAGASAIIGLLTS-----PLSKGJRGFALLESGLTR 450
 QY 509 -LADLEAOFAG-----LVNVYLIPDNGTGSMSLSKSPVPPAPAPLPQYTP 553
 DB 451 PLATLADSAASGERLADLSRLRSTDPATTLARADAPARSDLRPRPTSPYDGHVLP 510
 QY 554 QGPFPSVQGSVASSLMTFSEGLAFSGPRIFDVFOGERLYV-YEISIQ-----EALAIY 607
 DB 511 QTSDAIAIAGGLAVY--RVILGTINADEG-RAFLGARPMETRADYQAYLEAFGQGAALVA 567
 QY 608 -----GNSPAAMTTR-YVDGCFQMG-----KY-----TTPLTR 635
 DB 568 ACYPLDGRATPEKEMARIFGDNQFNRGVSAFSEALVROGAPVWRQFNGNTEGGRAPATH 627
 QY 636 GVDGCPYLATYVDHMFLESQAPKTRDAFCVFEQNO-GL-----PLRRHSDXLXSH 685
 DB 628 GAELIPYV-----FGVFKIDELDGLFDMPPGPTPADRALGQLMSS 666

QY 686 YFGGLAETLYVRSMSSTLNYDVMDFPHPSGATIRFYATGYISSAFLEGATGK 741
 DB 667 AMWRFAK-----NGDPAGDALTWPA-----YSTG--KSTMTFGEGGR 701

RESULT 11
 US-08-910-820-4
 ; Sequence 4, Application US/08910820
 ; Patent No. 6258579
 ; GENERAL INFORMATION:

; APPLICANT: Mercurio, Frank
 ; APPLICANT: Zhu, Hengyi
 ; APPLICANT: Barbosa, Miguel
 ; APPLICANT: Li, Gfan
 ; APPLICANT: Murray, Brian W.
 ; TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
 ; TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle

; STATE: Washington
 ; COUNTRY: USA

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/910,820

; FILING DATE: 12-AUG-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Makl, David J. 31,392

; REGISTRATION NUMBER: 860098,41361

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 272 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; US-08-910-820-4

Query Match 21.7%; Score 1162; DB 4; Length 272;
 Best Local Similarity 80.8%; Pred. No. 8.6e-106;
 Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGWKIKGLVOPTRILLLEYLEKEEYERDEGDKRNKKFELGLEFPNLPYYIDG 97
 DB 2 SPILGWKIKGLVOPTRILLLEYLEKEEYERDEGDKRNKKFELGLEFPNLPYYIDG 61
 QY 98 DVKLTQSMATIRYIADKNNMLGSGPKERAEISMLEGAVLDIRYGSRIAYSKDEFTLKYD 157
 DB 62 DVKLTQSMATIRYIADKNNMLGSGPKERAEISMLEGAVLDIRYGSRIAYSKDEFTLKYD 121
 QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDPMYDALDVYLYNDPCLDAFPKLVSEFK 217
 DB 122 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDPMYDALDVYLYNDPCLDAFPKLVSEFK 181
 QY 218 RIEAIPOIDKYLKSSKIYAMPLOGMOATFGGSDHPKSSQSEALFQGGDGEPSQLPHCP 277
 DB 182 RIEAIPOIDKYLKSSKIYAMPLOGMOATFGGSDHPKSSQSEALFQGGDGEPSQLPHCP 235
 QY 278 SVSPSAOPWTHPGOSOLFADLSREELTAVMRFLTQRLPGL 318
 DB 236 KGTADADEWCDSGLSLGPDAAAPG-----GPGL 264

RESULT 12
US-08-327-874A-11
Sequence 11, Application US/08327874A
Patent No. 6372249
GENERAL INFORMATION:
APPLICANT: BAYLOR COLLEGE OF MEDICINE
APPLICANT: SMITH, JAMES R.
APPLICANT: DRUTZ, DAVID J.
APPLICANT: WILSON, DEBORAH R.
APPLICANT: ZIMSTEIN, LOUIS A.
TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
TITLE OF INVENTION: DNA SYNTHESIS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS & WELLS
STREET: 200 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10166
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,874A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT US94/09700
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/327,874
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: NORTON, GERARD P.
REGISTRATION NUMBER: 36,621
REFERENCE/DOCKET NUMBER: 3634-8-CIP10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 878-3148
TELEFAX: (212) 878-8375
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
IMMEDIATE SOURCE:

CLONE: GST-SDI-1 fusion protein
US-08-327-874A-11
Query Match 21.7%; Score 1162; DB 4; Length 397;
Best Local Similarity 83.9%; Pred. No. 1.7e-105;
Matches 224; Conservative 10; Mismatches 29; Indels 4; Gaps 2;
QY 38 SPILGYKIKGLVOPTRLLLEYLEEKYEHLPERDECDKRNKKFELGEPNLPYIDG 97
DB 2 SPILGYKIKGLVOPTRLLLEYLEEKYEHLPERDECDKRNKKFELGEPNLPYIDG 61
QY 98 DVKLTGSMATIRYADKNNMGGSPKERAISMLEGALDIDRYGVSRIRYASKDFETLKYD 157
DB 62 DVKLTGSMATIRYADKNNMGGSPKERAISMLEGALDIDRYGVSRIRYASKDFETLKYD 121
QY 158 FLSTKPEMLKMFEDRLSHKTYLNGDHTYHPDPMILYDALDVLYLMDPKCLAFPKLYSFK 217
DB 122 FLSTKPEMLKMFEDRLCHKTYLNGDHTYHPDPMILYDALDVLYLMDPKCLAFPKLYSFK 181
QY 218 RIEAIPDIDKYLKSSKYIAMPLOGMQATFGGDDHPKPSQSQSEALFQGGDGEPSQLPHCP 277
DB 182 RIEAIPDIDKYLKSSKYIAMPLOGMQATFGGDDHPKPSQSQSEALFQGGDGEPSQLPHCP 277
QY 278 SVSPSAQPMWTHPGQSOLFADLSREELT 304
DB 238 AGDVQNPCCSKACRRFLGPDVSEQLS 264
RESULT 13
PCT-US94-09700-11
Sequence 11, Application PC/TUS9409700
GENERAL INFORMATION:
APPLICANT: BAYLOR COLLEGE OF MEDICINE
APPLICANT: SMITH, JAMES R.
TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
TITLE OF INVENTION: DNA SYNTHESIS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWREY & SIMON
STREET: 1299 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09700
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535

FILED DATE: 13-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: AUERBACH, JEFFREY I.
REGISTRATION NUMBER: 32,680
REFERENCE/DOCKET NUMBER: 225-102-C1P7-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 383-7451
TELEFAX: (202) 383-6610
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
IMMEDIATE SOURCE:
CLONE: GST-SDI-1 fusion protein
PCT-US94-09700-11

Query Match 21.7%; Score 1162; DB 5; Length 397;
Best Local Similarity 83.9%; Pred. No. 1.7e-105;
Matches 224; Conservative 10; Mismatches 29; Indels 4; Gaps 2;

QY 38 SPLIGTWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYIYG 97
DB 2 SPLIGTWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYIYG 61
QY 98 DVKLTOSMAIRIYADKHNMLGSGPKERAEISMLEGAVLDIRGVSRISAKDFETLKVD 157
DB 62 DVKLTOSMAIRIYADKHNMLGSGPKERAEISMLEGAVLDIRGVSRISAKDFETLKVD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHYTHDPFMYDALDVLYMDPCLDAPFLVCFKK 217
DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHYTHDPFMYDALDVLYMDPCLDAPFLVCFKK 181
QY 218 RIRAIPIIDKYLKSSKIAPLOGWMTFGGSDHPKPSQSQEALFOGDDGSEPSQLPHCP 277
DB 182 RIRAIPIIDKYLKSSKIAPLOGWMTFGGSDHPKPSQSQEALFOGDDGSEPSQLPHCP 237
QY 278 SVSPSAQPTWTHPGSQQLFADLSREELT 304
DB 238 AGDVQRNPGCKACRRLRFGPVDSQLS 264

RESULT 14
US-08-910-820-6
Sequence 6, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
State: Washington
Country: USA
Zip: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-6

Query Match 21.7%; Score 1161; DB 4; Length 272;
Best Local Similarity 80.8%; Pred. No. 1.1e-105;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPLIGTWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYIYG 97
DB 2 SPLIGTWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYIYG 61
QY 98 DVKLTOSMAIRIYADKHNMLGSGPKERAEISMLEGAVLDIRGVSRISAKDFETLKVD 157
DB 62 DVKLTOSMAIRIYADKHNMLGSGPKERAEISMLEGAVLDIRGVSRISAKDFETLKVD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHYTHDPFMYDALDVLYMDPCLDAPFLVCFKK 217
DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHYTHDPFMYDALDVLYMDPCLDAPFLVCFKK 181
QY 218 RIRAIPIIDKYLKSSKIAPLOGWMTFGGSDHPKPSQSQEALFOGDDGSEPSQLPHCP 277
DB 182 RIRAIPIIDKYLKSSKIAPLOGWMTFGGSDHPKPSQSQEALFOGDDGSEPSQLPHCP 235
QY 278 SVSPSAQPTWTHPGSQQLFADLSREELTAVRFLTORLPGI 318
DB 236 GKTADDEWCDAGLGPDAAPG-----GPGI 264

RESULT 15
US-08-506-2968-71
Sequence 71, Application US/08506296B
Patent No. 6313265
GENERAL INFORMATION:
APPLICANT: Phillips, Greg
APPLICANT: Cunningham, Bruce A.
APPLICANT: Crossin, Kathryn L.
TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute
STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
City: La Jolla
State: California
Country: U.S.
Zip: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506,296B
FILING DATE: 24-JUL-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937

TELEFAX: (619) 554-6312
 INFORMATION FOR SEQ ID NO: 71:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 635 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: C-terminal
 US-08-506-296B-71

Query Match 21.7%; Score 1161; DB 4; Length 635;
 Best Local Similarity 61.6%; Pred. No. 4.7e-105;
 Matches 252; Conservative 11; Mismatches 51; Indels 95; Gaps 11;

QY 38 SPILGWKIKGLVOPFRLLLEYLEEKEEHLERDEGDKMKNKKFELGLEFPNLPYIDG 97
 DB 2 SPILGWKIKGLVOPFRLLLEYLEEKEEHLERDEGDKMKNKKFELGLEFPNLPYIDG 61
 QY 98 DVKLQSMALIRIADKHNMLGSGPKERAEISMLEGAVLDIRYGVSRIRAYSKDFETLKVD 157
 DB 62 DVKLQSMALIRIADKHNMLGSGPKERAEISMLEGAVLDIRYGVSRIRAYSKDFETLKVD 121
 QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPMCLDAFPKLVSKK 217
 DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYMDPMCLDAFPKLVSKK 181
 QY 218 RIRAIPOIDKYLKSKYIAMPLOGMATFEGGDHPKKS-----OSQE 259
 DB 182 RIRAIPOIDKYLKSKYIAMPLOGMATFEGGDHPKSDLVPRGSHIHKSHIVVPANTTS 241
 QY 260 ALFG-----GDGGEPSQLP--HCPVSPSAQPWT 287
 DB 242 AILSGLRPYSSYHVEVQAFNGRGLGRASEWTFSTPEGVPGHPFLHLECOSDTSLLHMQ 301
 QY 288 -----HP--GQS--QLFADLSREELTAVMRFLTQRLGPGLVDAQAQAPS 327
 DB 302 PPLSHNGVLTYGLSYHPVEGESKEOLFNLSDPELRT--HNLV--NLNPD----- 349
 QY 328 DNCVFSEVLQPPKAALAHLDGSPPPA--REALAVFFGRQPPQPVYS 374
 DB 350 ---QYRFQLO-----ATTQGGPQQAIVREGTVALFGKPDDEGNIS 387

Search completed: May 20, 2003, 11:41:03
 Job time : 37.1643 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 159.01 Seconds

(without alignments)
622.443 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357

Sequence: 1 MDWLRLFLFMAAOSINAA.....QAACAPDLPAFSGGFSHN 998

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5357	100.0	998	9 US-10-081-408-20	Sequence 20, Appl
2	3667	74.1	763	9 US-10-081-408-2	Sequence 2, Appl
3	2480.5	46.3	729	10 US-09-919-497-51	Sequence 51, Appl
4	1165.5	21.8	401	9 US-10-190-866A-1	Sequence 1, Appl
5	1162	21.7	272	9 US-09-844-988-4	Sequence 4, Appl
6	1162	21.7	272	10 US-09-844-908-4	Sequence 4, Appl
7	1161	21.7	272	9 US-09-844-988-6	Sequence 6, Appl
8	1161	21.7	272	10 US-09-844-908-6	Sequence 6, Appl
9	1160.5	21.7	298	9 US-09-844-908-6	Sequence 6, Appl
10	1159	21.6	218	9 US-09-910-600-26	Sequence 26, Appl
11	1157.5	21.6	422	9 US-09-824-438-7	Sequence 5, Appl
12	1157	21.6	331	9 US-09-824-438-6	Sequence 7, Appl
13	1155.5	21.6	1140	9 US-09-950-634-4	Sequence 6, Appl
14	1154	21.5	282	9 US-09-844-988-3	Sequence 4, Appl
15	1154	21.5	282	9 US-09-844-988-5	Sequence 3, Appl
16	1154	21.5	282	10 US-09-844-908-3	Sequence 3, Appl
17	1154	21.5	282	10 US-09-844-908-5	Sequence 5, Appl
18	1153	21.5	412	9 US-09-775-964-34	Sequence 34, Appl
19	1152	21.5	364	9 US-09-788-268-14	Sequence 14, Appl

20	1151	21.5	239	10 US-09-823-153-7	Sequence 7, Appl
21	1150	21.5	245	9 US-10-267-311-23	Sequence 23, Appl
22	1150	21.5	394	10 US-09-990-578-4	Sequence 4, Appl
23	1148.5	21.4	354	10 US-09-823-153-8	Sequence 8, Appl
24	1148	21.4	374	9 US-10-267-311-25	Sequence 25, Appl
25	1148	21.4	348	9 US-09-810-600-22	Sequence 22, Appl
26	1148	21.4	348	9 US-09-910-600-23	Sequence 23, Appl
27	1148	21.4	348	9 US-09-910-600-24	Sequence 24, Appl
28	1148	21.4	348	9 US-09-910-600-25	Sequence 25, Appl
29	1145	21.4	649	9 US-10-174-784-9	Sequence 9, Appl
30	1144	21.4	218	9 US-10-081-408-4	Sequence 4, Appl
31	482.5	9.0	229	9 US-10-102-806-554	Sequence 554, App
32	196.5	3.7	222	9 US-09-784-739-4	Sequence 4, Appl
33	196.5	3.7	220	10 US-09-925-301-1318	Sequence 1318, Ap
34	179.5	3.4	222	9 US-09-784-739-3	Sequence 3, Appl
35	173.5	3.2	222	9 US-09-784-739-5	Sequence 5, Appl
36	169.5	3.2	203	9 US-09-847-208-48	Sequence 48, Appl
37	164.5	3.1	222	9 US-09-784-739-1	Sequence 1, Appl
38	109.5	2.0	240	9 US-09-864-899-23	Sequence 23, Appl
39	106	2.0	233	10 US-09-765-213A-2	Sequence 2, Appl
40	106	2.0	617	9 US-10-005-530-58	Sequence 58, Appl
41	104	1.9	980	9 US-09-908-193-22	Sequence 22, Appl
42	103	1.9	2224	9 US-10-115-563-14	Sequence 14, Appl
43	102	1.9	1194	9 US-09-738-626-5832	Sequence 5832, Ap
44	101	1.9	617	9 US-10-005-530-46	Sequence 46, Appl
45	99	1.8	233	10 US-09-765-213A-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-10-081-408-20
; Sequence 20, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abtams, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE NUMBER: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081.408
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 010625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 998
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant construct
US-10-081-408-20

Query Match      100.0%; Score 5357; DB 9; Length 998;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 998; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	MDWLRLFLFMAAOSINAAOHDAVNRKKEQONASPIIGYKIKGLVOPTRLLEYL	60
DB	1	MDWLRLFLFMAAOSINAAOHDAVNRKKEQONASPIIGYKIKGLVOPTRLLEYL	60
QY	61	EEKEEHLVLRDEGDKRNKKEFLEPNLPYIIDGVKLTOSMAIIRYADHNMGG	120
DB	61	EEKEEHLVLRDEGDKRNKKEFLEPNLPYIIDGVKLTOSMAIIRYADHNMGG	120
QY	121	SPKRRADISMLEGAVLDIRGVSRIRYASKDETAKVDFLSKLPMLKMFEDRLSHKTYLN	180
DB	121	SPKRRADISMLEGAVLDIRGVSRIRYASKDETAKVDFLSKLPMLKMFEDRLSHKTYLN	180
QY	181	GDHYTHDFMLYDALDVYLLTMDPCLDAPFLVLSFKKRIEAIPOIDIKYTLASSTIANPQ	240

```

Db 181 GDHTHDFHLYDLVDLVLMWDPKCLAPFLVSEFKRIEALPQIDKYLKSSKIAMPLO 240
QY 241 GWOATFGGCHPRKPSQSQEALFQGGDGEPSQLPHCPSPVSAQPTWHPGOSQFLADLSR 300
Db 241 GWOATFGGCHPRKPSQSQEALFQGGDGEPSQLPHCPSPVSAQPTWHPGOSQFLADLSR 300
QY 301 EELTAVNRFLTORGLPGLVYAAOARPSDNCVFSEYELDLPRAAALAHLDKRGSPAREAL 360
Db 301 EELTAVNRFLTORGLPGLVYAAOARPSDNCVFSEYELDLPRAAALAHLDKRGSPAREAL 360
QY 361 AIYFFGROPVNSSELYVGLPHPSYRDYVERHGGPLPYHRRPVLFOEYLDIDOMIFN 420
Db 361 AIYFFGROPVNSSELYVGLPHPSYRDYVERHGGPLPYHRRPVLFOEYLDIDOMIFN 420
QY 421 RELPQASGLLHCCFYKHKRGNLYTTTAPRGLOSGBRATFGLYINISGAGFLHNG 480
Db 421 RELPQASGLLHCCFYKHKRGNLYTTTAPRGLOSGBRATFGLYINISGAGFLHNG 480
QY 481 ELLVNHKALDPAKWTIOKVFQGRYDSLAOLEAOFEGALVNVVLPDNGTGGWSLSP 540
Db 481 ELLVNHKALDPAKWTIOKVFQGRYDSLAOLEAOFEGALVNVVLPDNGTGGWSLSP 540
QY 541 VPPGAPPLQFYPOGPRFSVSGSRVASSLMTFSGLAFSGPRIFDVRFQGERLYEISL 600
Db 541 VPPGAPPLQFYPOGPRFSVSGSRVASSLMTFSGLAFSGPRIFDVRFQGERLYEISL 600
QY 601 QEALATIGGNSPAAATRYVDGSGMKYTTPLTRGVDCPLATYVDVHFLLESQAPRTI 660
Db 601 QEALATIGGNSPAAATRYVDGSGMKYTTPLTRGVDCPLATYVDVHFLLESQAPRTI 660
QY 661 RDAFCVEQONGPLRRHNSDLYHYFGLAEYLVVRSMTLLNDYVMDPTVFHPSGAI 720
Db 661 RDAFCVEQONGPLRRHNSDLYHYFGLAEYLVVRSMTLLNDYVMDPTVFHPSGAI 720
QY 721 EIRFYATGYISSAFLEFGATGKYGNQVSEHTLGTVHTSHAKFKVDLDVAGLENNVMAEDM 780
Db 721 EIRFYATGYISSAFLEFGATGKYGNQVSEHTLGTVHTSHAKFKVDLDVAGLENNVMAEDM 780
QY 781 FVPAVWPSPHOLQLOVTRKLLMEEOAFLVGSATPRLYTLASNSNMKGHRGRI 840
Db 781 FVPAVWPSPHOLQLOVTRKLLMEEOAFLVGSATPRLYTLASNSNMKGHRGRI 840
QY 841 QMLSFAGEPLPONSMMARGFSMERYQLAVTORKEEPPSSSVFNQNDMPATVDSDFIN 900
Db 841 QMLSFAGEPLPONSMMARGFSMERYQLAVTORKEEPPSSSVFNQNDMPATVDSDFIN 900
QY 901 NETTAGKDLVAVTAGFLHIHAEIDPNTVTVGNGVGFLLRPYNFDEDPSTYSADSIYF 960
Db 901 NETTAGKDLVAVTAGFLHIHAEIDPNTVTVGNGVGFLLRPYNFDEDPSTYSADSIYF 960
QY 961 RGDODAGACEVNPPLACLPOAAACAPDLPAFSGGFSHN 998
Db 961 RGDODAGACEVNPPLACLPOAAACAPDLPAFSGGFSHN 998

```

RESULT 2

US-10-081-408-2

Sequence 2, Application US/10081408

Patent No. US20020160482A1

GENERAL INFORMATION:

APPLICANT: Abrahams, n, Lars

TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION

FILE REFERENCE: 13425-053001

CURRENT APPLICATION NUMBER: US/10/081.408

PRIOR FILING DATE: 2002-02-21

PRIOR APPLICATION NUMBER: SE 0100625-3

PRIOR FILING DATE: 2001-02-23

PRIOR APPLICATION NUMBER: US 60/272.247

NUMBER OF SEQ ID NOS: 20

SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 2
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-081-408-2

```

```

Query Match          74.1%; Score 3967; DB 9; Length 763;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 735; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 263 QGGDGEPSQLPHCPSPVSAQPTWHPGOSQFLADLSREELTAVNRFLTORGLGGLYDAA 322
Db 28 RGGDGEPSQLPHCPSPVSAQPTWHPGOSQFLADLSREELTAVNRFLTORGLGGLYDAA 87
QY 323 QARSDNCVFSEYELDLPRAAALAHLDKRGSPAREALATYFFGROPVNSSELYVGLP 382
Db 88 QARSDNCVFSEYELDLPRAAALAHLDKRGSPAREALATYFFGROPVNSSELYVGLP 147
QY 383 HPSYRDYVERHGGPLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHCCFYKHKRGN 442
Db 148 HPSYRDYVERHGGPLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHCCFYKHKRGN 207
QY 443 LVTMTAPRGLOSGBRATFGLYINISGAGFLHNGVLELLVNHKALDPAKWTIOKVFYQ 502
Db 208 LVTMTAPRGLOSGBRATFGLYINISGAGFLHNGVLELLVNHKALDPAKWTIOKVFYQ 267
QY 503 GRYYDSLAOLEAOFEGALVNVVLPDNGTGGWSLSPVPPGAPPLQFYPOGPRFSYOG 562
Db 268 GRYYDSLAOLEAOFEGALVNVVLPDNGTGGWSLSPVPPGAPPLQFYPOGPRFSYOG 327
QY 563 SRVASSLMTFSGLAFSGPRIFDVRFQGERLYEISLOEALATYGNSSPAAATTRYVDG 622
Db 328 SRVASSLMTFSGLAFSGPRIFDVRFQGERLYEISLOEALATYGNSSPAAATTRYVDG 387
QY 623 GFGMKYTTPLTRGVDCYLATYVDVHFLLESQAKTIRDAFCVEQONGPLRRHNSDL 682
Db 388 GFGMKYTTPLTRGVDCYLATYVDVHFLLESQAKTIRDAFCVEQONGPLRRHNSDL 447
QY 683 YSHYFGLAEYLVVRSMTLLNDYVMDTVFHPGSAIEIRFYATGYISSAFLEFGATGY 742
Db 448 YSHYFGLAEYLVVRSMTLLNDYVMDTVFHPGSAIEIRFYATGYISSAFLEFGATGY 507
QY 743 GNQVSEHTLGTVHTSHAKFKVDLDVAGLENNVMAEDMVFVPAVWPSPHOLQLOVTRK 802
Db 508 GNQVSEHTLGTVHTSHAKFKVDLDVAGLENNVMAEDMVFVPAVWPSPHOLQLOVTRK 567
QY 803 LLEMEEOAFLVGSATPRLYTLASNSNMKGHRGRIOMLSFAGEPLPONSMMARGFSM 862
Db 568 LLEMEEOAFLVGSATPRLYTLASNSNMKGHRGRIOMLSFAGEPLPONSMMARGFSM 627
QY 863 ERYQLAVTORKEEPPSSSVFNQNDMPATVDSFINNETTAGKDLVAVTAGFLHIH 922
Db 628 ERYQLAVTORKEEPPSSSVFNQNDMPATVDSFINNETTAGKDLVAVTAGFLHIH 687
QY 923 AEDIPNTVTVGNGVGFLLRPYNFDEDPSTYSADSIYFRGDODAGACEVNPPLACLPOAAA 982
Db 688 AEDIPNTVTVGNGVGFLLRPYNFDEDPSTYSADSIYFRGDODAGACEVNPPLACLPOAAA 747
QY 983 CAPDLPAFSGGFSHN 998
Db 748 CAPDLPAFSGGFSHN 763

```

RESULT 3

US-09-919-497-51

Sequence 51, Application US/09919497

Patent No. US20020106662A1

GENERAL INFORMATION:

APPLICANT: Mutter, George L.

TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER

CURRENT APPLICATION NUMBER: US/09/919.497

CURRENT FILING DATE: 2001-07-31

; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-51

Query Match 46.3%; Score 2480.5; DB 10; Length 729;
Best Local Similarity 64.1%; Pred. No. 2,4e-203;
Matches 468; Conservative 86; Mismatches 145; Indels 31; Gaps 3;

Qy 268 GEPQLPHCPSPSPASOPTHPGOSOLFADLSREELTAVMRFOTRLGGLVDAQAQPS 327
Db 29 GSSQPPHCPSPVSHRQPPHPCOSQLFADLSREELTAVMRFOTRLGGLVDAQAQPS 88
Qy 328 DNCVPSVELQDPPKAAALAHLDGSPPARALAIYFSGROPVNSVGLVGPPLPSPYM 387
Db 89 DNCISFVELQDPPKAAALAHLDGSPPARALAIYFSGROPVNSVGLVGPPLPSPYM 148
Qy 388 RQVTERHGGCLPYHRRPPLFOEYDIDQMIFRRELPOASGLHCCCFKHHGRNLVTMT 447
Db 149 RQVTERHGGCLPYHRRPPLFOEYDIDQMIFRRELPOASGLHCCCFKHHGRNLVTMT 206
Qy 448 TAPRGISGDRATFGLYVNSGAGFFLHHVGLLELVNKKALDPAWITQKVEYQGRYD 507
Db 207 ATRPGIRSRERTWITGLYNNISVGLFHPVGLLELLDRAIDPAWITQKVEYQGRYD 266
Qy 508 SLAQLEAOFAGLVNVLIPDNGTSGWSLKSFPVPPAPLPLOFYQGRFVSQGRVAS 567
Db 267 DLGQLEERKSGRLLEVRYVPLPPNGASSLSRNSPGLPPLQFSPGSOYVQGNLVYS 326
Qy 568 SLWTFSEFGAGSPRIFFVREGGERLYEISLOEALATYGGNSPAMTTRVYDGGFGMG 627
Db 327 SLWTFSEFGAGVSGRLIFDVRFOGERIATEVSQVCSYTAGDSPTMLTRYDSSFGDG 386
Qy 628 KYTTPLRGVDPCYATYVDMHFLLESQAPRTIRDAFCVEEONQGLPLRRHNSDLXSHYF 687
Db 387 RNSRGVLRGVDCYQATYVDMHFLVKGAVQLLPGAVCFEEAQLPLRRHNVLDNHVF 446
Qy 688 GGIAGELVLYRKSSTLNTDYWDYFHPSGAIEIRPYATGYSSAFLEGATG--KYGNQ 745
Db 447 GGIAGELVLYRKSSTLNTDYWDYFHPSGAIEIRPYATGYSSAFLEGATG--KYGNQ 506
Qy 746 VSEHTIGTVHTSHAFKVDLVAGLEMMVAEDMVFVMAVPWSPREHOLQRLQVTRKLE 805
Db 507 VGERVIGTVHTSHAFKVDLVAGLEMMVAEDMVFVMAVPWSPREHOLQRLQVTRKLE 566
Qy 806 MEQOAAFLVGSATPRYLYLASNSHKNKGHPRGYRIQMLSFAGEPLPQNSSMARGFSWERY 865
Db 567 KEDLTAFSLGSPPLRYLYLASNOTNAMGHORG-----Y 599
Qy 866 QLAVTQRKEEPSSSVFNQNDPMAPRYVDSDFINNETINGKDLVAVYTAGFLHIAED 925
Db 600 QLAVTQRKEEPSSSVFNQNDPMAPRYVDSDFINNETINGKDLVAVYTAGFLHIAED 659
Qy 926 IPMTVTVGNGVGFLLRPYNFDEDPSPFYSADSIYFRGDODAGACEVPLCLPOAAACAP 985
Db 660 IPMTVTVGNGVGFLLRPYNFDEDPSPFYSADSIYFRGDODAGACEVPLCLPOAAACAP 719
Qy 986 DLPFASHGCF 995
Db 720 DLPFASHGCF 729

RESULT 4
US-10-190-866A-1
; Sequence 1, Application US/10190866A
; Publication No. US20030092619A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim International GmbH

; TITLE OF INVENTION: Method for identifying substances which modulate
; TITLE OF INVENTION: Interleukin 4 signalling
; FILE REFERENCE: 1/1214 US
; CURRENT APPLICATION NUMBER: US/10/190,866A
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 60/306,552
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: EP 01117097.4
; PRIOR FILING DATE: 2001-07-13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: derivative of
; OTHER INFORMATION: STAT6 fragment
US-10-190-866A-1

Query Match 21.8%; Score 1165.5; DB 9; Length 401;
Best Local Similarity 68.7%; Pred. No. 3,4e-91;
Matches 248; Conservative 10; Mismatches 56; Indels 47; Gaps 10;

Qy 38 SPILGYWKIKGLVQPTRLLELYEKEEYERDEGDKWRNKKFELGLEFPNLPYIDG 97
Db 2 SPILGYWKIKGLVQPTRLLELYEKEEYERDEGDKWRNKKFELGLEFPNLPYIDG 61
Qy 98 DVKLTQSMATIRYIADKNNMGLGSPKERAISMLBGAVIDIRGVSRIYNSKQFELKYD 157
Db 62 DVKLTQSMATIRYIADKNNMGLGSPKERAISMLBGAVIDIRGVSRIYNSKQFELKYD 121
Qy 158 FLTKLEPMKMPEDRLSHKTYLNGDVHTPDVFLYDALDVLYLMDPCIDAPPKLYSFKK 217
Db 122 FLTKLEPMKMPEDRLCHNTYLDGHTVHPDFMLYDALDVLYLMDPCIDAPPKLYSFKK 181
Qy 218 RIFALPQIDKYLYKSSYTIAMPLOGWATFEGGDPKPSOSQELFQGG-----DGEFPGQ 272
Db 182 RIFALPQIDKYLYKSSYTIAMPLOGWATFEGGDPKPSOSQELFQGG-----DGEFPGQ 237
Qy 273 LPHCPSPVSAOPTHPGOSOLFADLSREELTAVMRFOTRLGGLVDAQAQPSDNCVF 332
Db 238 LG--PDMVPOVYF-PHSHSIPPIQGLSPEESVAVL-----SAPQEP----- 275
Qy 333 SVELQDPPKAAALAHLDGSPPARALAIYFSGROPVNSVGLVGPPLPSPYM 391
Db 276 --HLQPPSLGQWS-LPFDQPHFG-----LLPCQGOEHAVSS-----PDLCLSDVT 320
Qy 392 V 392
Db 321 M 321

RESULT 5
US-09-844-988-4
; Sequence 4, Application US/09844988
; Patent No. US20020158764A1
; GENERAL INFORMATION:
; APPLICANT: Mercurio, Frank
; Zhu, Hengyi
; Barbosa, Miguel
; Li, Gian
; Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,988
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/910,820
FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-988-4

Query Match 21.7%; Score 1162; DB 9; Length 272;
Best Local Similarity 80.8%; Pred. No. 3.6e-91;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGYWKIKGLVQPTLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
DB 2 SPILGYWKIKGLVQPTLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
QY 98 DVKLTSMAIIRYIADKHNMLGGSPKERATISMEGAVLDIRYGVSRIVASKDFETLKVD 157
DB 62 DVKLTSMAIIRYIADKHNMLGGSPKERATISMEGAVLDIRYGVSRIVASKDFETLKVD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLYMDPCLDAFPKLVCFKK 217
DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCLDAFPKLVCFKK 181
QY 218 RIBAIPOIDKYLKSKXYIAMPLOGWQATFGGDHPKPSQSQSEALFOGGDGSEPSQLPHCP 277
DB 182 RIBAIPOIDKYLKSKXYIAMPLOGWQATFGGDHPKPSDPRREFIVTDMAG-----VACL 235
QY 278 SVSPSAQPTWTHPGQSQLEFADLSREELTAVNRFLTQRLGPG 318
DB 236 GKTADADEMCDSGLSGLPDAAPG-----GPG 264

RESULT 6

US-09-844-908-4
Sequence 4, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gian

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-Aug-1997

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-908-4

Query Match 21.7%; Score 1162; DB 10; Length 272;
Best Local Similarity 80.8%; Pred. No. 3.6e-91;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGYWKIKGLVQPTLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
DB 2 SPILGYWKIKGLVQPTLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
QY 98 DVKLTSMAIIRYIADKHNMLGGSPKERATISMEGAVLDIRYGVSRIVASKDFETLKVD 157
DB 62 DVKLTSMAIIRYIADKHNMLGGSPKERATISMEGAVLDIRYGVSRIVASKDFETLKVD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLYMDPCLDAFPKLVCFKK 217
DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCLDAFPKLVCFKK 181
QY 218 RIBAIPOIDKYLKSKXYIAMPLOGWQATFGGDHPKPSQSQSEALFOGGDGSEPSQLPHCP 277
DB 182 RIBAIPOIDKYLKSKXYIAMPLOGWQATFGGDHPKPSDPRREFIVTDMAG-----VACL 235
QY 278 SVSPSAQPTWTHPGQSQLEFADLSREELTAVNRFLTQRLGPG 318
DB 236 GKTADADEMCDSGLSGLPDAAPG-----GPG 264

RESULT 7

US-09-844-988-6
Sequence 6, Application US/09844988
Patent No. US20020158764A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gian

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,988
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/910,820
FILING DATE: 1997-08-13
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413c1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-844-988-6

Query Match      21.7%; Score 1161; DB 9; Length 272;
Best Local Similarity 80.8%; Pred. No. 4.4e-91;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGWKIKGLVQPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
DB 2 SPILGWKIKGLVQPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
QY 98 DVKLTOSMAIIRYIADKHNLMGSPKRAEISMEGAVLDIRYVSNIAVSKDETLKVD 157
DB 62 DVKLTOSMAIIRYIADKHNLMGSPKRAEISMEGAVLDIRYVSNIAVSKDETLKVD 121
QY 158 FLKSLPMLKMFEDRLSHKTYLNGDHTHPDPMLYDALDVLVYNDPMCLDAFPPLVSEFK 217
DB 122 FLKSLPMLKMFEDRLSHKTYLNGDHTHPDPMLYDALDVLVYNDPMCLDAFPPLVSEFK 181
QY 218 RIEAIPQIDKYLKSSKXIYAMPLOGMQATFGGSDHPKPSQSQSEALFQGGDGEPSQLPHCP 277
DB 182 RIEAIPQIDKYLKSSKXIYAMPLOGMQATFGGSDHPKPSQSQSEALFQGGDGEPSQLPHCP 235
QY 278 SVSPSAQPWTHPGQSOLFADLSREELTAVMFLTQRLGPG 318
DB 236 GKTADADEMCDAGLGAAGPDAAPG-----GPG 264

RESULT 8
US-09-844-908-6
Sequence 6, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercutio, Frank
Zhu, Henry
Bardosa, Miguel
Li, Gian
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-Aug-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413c1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-844-908-6

Query Match      21.7%; Score 1161; DB 10; Length 272;
Best Local Similarity 80.8%; Pred. No. 4.4e-91;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGWKIKGLVQPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
DB 2 SPILGWKIKGLVQPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
QY 98 DVKLTOSMAIIRYIADKHNLMGSPKRAEISMEGAVLDIRYVSNIAVSKDETLKVD 157
DB 62 DVKLTOSMAIIRYIADKHNLMGSPKRAEISMEGAVLDIRYVSNIAVSKDETLKVD 121
QY 158 FLKSLPMLKMFEDRLSHKTYLNGDHTHPDPMLYDALDVLVYNDPMCLDAFPPLVSEFK 217
DB 122 FLKSLPMLKMFEDRLSHKTYLNGDHTHPDPMLYDALDVLVYNDPMCLDAFPPLVSEFK 181
QY 218 RIEAIPQIDKYLKSSKXIYAMPLOGMQATFGGSDHPKPSQSQSEALFQGGDGEPSQLPHCP 277
DB 182 RIEAIPQIDKYLKSSKXIYAMPLOGMQATFGGSDHPKPSQSQSEALFQGGDGEPSQLPHCP 235
QY 278 SVSPSAQPWTHPGQSOLFADLSREELTAVMFLTQRLGPG 318
DB 236 GKTADADEMCDAGLGAAGPDAAPG-----GPG 264

RESULT 9
US-09-910-600-26
Sequence 26, Application US/09910600
Publication No. US20030036631A1
GENERAL INFORMATION:
APPLICANT: Longphre, Malinda
Chang, Han
APPLICANT: Whitney, Gene
TITLE OF INVENTION: NOVEL SILENCES AND USES THEREOF
FILE REFERENCE: D0003NP
CURRENT APPLICATION NUMBER: US/09/910,600
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: 60/220,139
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 32
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 26
LENGTH: 298
TYPE: PRT
ORGANISM: Artificial Sequence
```


TITLE OF INVENTION: Beta-lipotropin and Uses Thereof

FILE REFERENCE: X-12138
CURRENT APPLICATION NUMBER: US/09/824,438
CURRENT FILING DATE: 2001-10-29
NUMBER OF SEQ ID NOS: 35
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO: 6
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GST fusion
OTHER INFORMATION: partner
US-09-824-438-6

Query Match 21.6%; Score 1157; DB 9; Length 331;
Best Local Similarity 72.1%; Pred. No. 1,3e-90;
Matches 235; Conservative 8; Mismatches 43; Indels 40; Gaps 5;

QY 38 SPLIGWTKIKGLVQPTRLLEYLEEYEEHLYRDEGDKWRNKKFELGLEFPNLPYIIDG 97
DB 2 SPLIGWTKIKGLVQPTRLLEYLEEYEEHLYRDEGDKWRNKKFELGLEFPNLPYIIDG 61
QY 98 DVKLTOSMAIIRYIADKHNMLGGSPKERAEISMLEGAVLDIRGVSRIVASKDFETLKVD 157
DB 62 DVKLTOSMAIIRYIADKHNMLGGSPKERAEISMLEGAVLDIRGVSRIVASKDFETLKVD 121
QY 158 FLTKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYLMDPCLDAFRLKYSFKK 217
DB 122 FLTKLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLYLMDPCLDAFRLKYSFKK 181
QY 218 RIFAIPQIDKYLKSKYTIAMPLOGMOTFGGSDHPKSSQSEALFQGGDGGESQLPKPC 217
DB 182 RIFAIPQIDKYLKSKYTIAMPLOGMOTFGGSDHPKSSQSEALFQGGDGGESQLPKPC 236
QY 278 SVSPSAQPTWHPGSQLFADLSREELTAVMRFLTORLGPVLDAQAARPDNCFVSVEIQ 337
DB 237 -----GSIEGRELTC-----QRLREG--DGPGPADDGAGQAADLE 270
QY 338 LPPKAAA-----LAHLDRGSP 354
DB 271 HSLVAERKDEGPRMEHFRMGSP 296

RESULT 13

US-09-950-634-4
Sequence 4, Application US/09950634
Publication No. US20030032775A1
GENERAL INFORMATION:

APPLICANT: Molnar-Kimber, Katherine L.
Fallili, Amedeo F.
Cagliano, Thomas J.
Nakanishi, Koji
Chen, Yanqiu

TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunne, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/950,634
FILING DATE: 13-Sep-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/471,112
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: US 08/384,524
FILING DATE: 13-FEB-1995
APPLICATION NUMBER: US 08/312,023
FILING DATE: 26-SEP-1995
APPLICATION NUMBER: US 08/207,975
FILING DATE: 08-MAR-1994

ATTORNEY/AGENT INFORMATION:
NAME: Slekmann, Michael T.
REGISTRATION NUMBER: 36,276
REFERENCE/DOCKET NUMBER: 01142.0058-00000

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 1140 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-950-634-4

Query Match 21.6%; Score 1155.5; DB 9; Length 1140;
Best Local Similarity 92.8%; Pred. No. 1,3e-89;
Matches 219; Conservative 1; Mismatches 11; Indels 5; Gaps 1;

QY 38 SPLIGWTKIKGLVQPTRLLEYLEEYEEHLYRDEGDKWRNKKFELGLEFPNLPYIIDG 97
DB 2 SPLIGWTKIKGLVQPTRLLEYLEEYEEHLYRDEGDKWRNKKFELGLEFPNLPYIIDG 61
QY 98 DVKLTOSMAIIRYIADKHNMLGGSPKERAEISMLEGAVLDIRGVSRIVASKDFETLKVD 157
DB 62 DVKLTOSMAIIRYIADKHNMLGGSPKERAEISMLEGAVLDIRGVSRIVASKDFETLKVD 121
QY 158 FLTKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYLMDPCLDAFRLKYSFKK 217
DB 122 FLTKLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLYLMDPCLDAFRLKYSFKK 181
QY 218 RIFAIPQIDKYLKSKYTIAMPLOGMOTFGGSDHPKSSQSEALFQGGDGGESQLPKPC 268
DB 182 RIFAIPQIDKYLKSKYTIAMPLOGMOTFGGSDHPKSSQSEALFQGGDGGESQLPKPC 237

RESULT 14

US-09-844-988-3
Sequence 3, Application US/09844988
Patent No. US20020158764A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Glen

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

1      APPLICATION NUMBER: US/09/844,988
2      FILING DATE: 26-Apr-2001
3      CLASSIFICATION: <Unknown>
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: 08/910,820
6      FILING DATE: 1997-08-13
7      ATTORNEY/AGENT INFORMATION:
8      NAME: Makl, David J.
9      REGISTRATION NUMBER: 31,392
10     REFERENCE/DOCKET NUMBER: 860098.413C1
11     TELECOMMUNICATION INFORMATION:
12     TELEPHONE: (206) 622-4900
13     TELEFAX: (206) 682-6031
14     INFORMATION FOR SEQ ID NO: 3:
15     SEQUENCE CHARACTERISTICS:
16     LENGTH: 283 amino acids
17     TYPE: amino acid
18     STRANDEDNESS: <Unknown>
19     TOPOLOGY: linear
20     SEQUENCE DESCRIPTION: SEQ ID NO: 3:
21     US-09-844-988-3

```

```

1      APPLICATION NUMBER: 08/910, 820
2      FILING DATE: 1997-08-13
3
4      ATTORNEY/AGENT INFORMATION:
5
6      NAME: Makl, David J.
7      REGISTRATION NUMBER: 31,392
8      REFERENCE/DOCKET NUMBER: 860098.413C1
9      TELECOMMUNICATION INFORMATION:
10     TELEPHONE: (206) 622-4900
11     TELEFAX: (206) 682-6031
12
13     INFORMATION FOR SEQ ID NO: 5:
14
15     SEQUENCE CHARACTERISTICS:
16
17     LENGTH: 282 amino acids
18     TYPE: amino acid
19     STRANDEDNESS: <Unknown>
20     TOPOLOGY: linear
21
22     SEQUENCE DESCRIPTION: SEQ ID NO: 5:
23     US-09-844-988-5

```

RESULT 15
 US-09-844-988-5
 Sequence 5, Application US/09844988
 Patent No. US20020158764A1
 GENERAL INFORMATION:
 APPLICANT: Mercurio, Frank
 Zhu, Hengyi
 Barbosa, Miguel
 Li, Gfan
 Murray, Brian W.
 TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
 COMPLEX AND METHODS OF USE THEREFOR
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/844,988
 FILING DATE: 26-Apr-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:09:54 ; Search time 38.5066 Seconds
(without alignments)
2491.578 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357
Sequence: 1 MDWLRNLLFLMAAQAQINMA.....QAAACAPDLPAFSGHGFSSN 998

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73: *
2: PIR1: *
3: PIR2: *
4: PIR3: *
5: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3967	74.1	763	2 JC5234	amine oxidase (cop
2	3326	62.1	762	2 A54411	amine oxidase (cop
3	1398.5	26.1	746	2 S34656	amine oxidase (cop
4	1394	26.0	751	2 A54053	amine oxidase (cop
5	1140	21.3	219	2 A26484	glutathione transf
6	985	18.4	218	2 A45556	glutathione S-tran
7	922	17.2	218	2 A45523	glutathione transf
8	663	12.4	209	2 A48388	glutathione S-tran
9	489.5	9.1	218	2 S33860	glutathione transf
10	489.5	9.1	218	2 A29794	glutathione transf
11	486.5	9.1	218	2 S32425	glutathione transf
12	485.5	9.1	218	2 S36574	glutathione transf
13	484.5	9.0	218	2 A39375	glutathione transf
14	483.5	9.0	218	2 A47486	glutathione transf
15	481.5	9.0	217	2 JX0095	glutathione transf
16	476.5	8.9	218	2 S13202	glutathione transf
17	476.5	8.9	218	2 A23732	glutathione transf
18	475.5	8.9	218	2 B34159	glutathione transf
19	473.5	8.8	218	2 S01719	glutathione transf
20	471.5	8.8	218	2 A46019	glutathione transf
21	469.5	8.8	218	2 A29036	glutathione transf
22	468.5	8.7	218	2 B29231	glutathione transf
23	463.5	8.7	218	2 XURG4	glutathione transf
24	460.5	8.6	218	2 A46143	mu-class glutathio
25	452.5	8.4	225	2 A35295	glutathione transf
26	437.5	8.2	220	2 S18464	glutathione transf
27	413	7.7	735	2 B41836	amine oxidase (fla
28	397.5	7.4	757	2 E64889	amine oxidase (cop

30	395.5	7.4	219	2 S50146	major allergen Dp1
31	377	7.0	684	2 A56102	amine oxidase (cop
32	368.5	6.9	660	2 G90330	amine oxidase (cop
33	363	6.8	648	2 A48646	amine oxidase (cop
34	357.5	6.7	756	2 T48139	copper amine oxida
35	353.5	6.6	650	2 G71412	probable amine oxi
36	340	6.3	638	2 JC2139	phenylethylamine o
37	325.5	6.1	674	2 JC7251	amine oxidase (cop
38	324	6.0	687	2 T47403	amine oxidase-like
39	320	6.0	674	2 C44239	amine oxidase (cop
40	305	5.7	759	2 E84854	probable copper am
41	300	5.6	660	2 AH2234	copper amine oxida
42	295	5.5	692	2 S04963	amine oxidase (cop
43	260.5	4.9	142	2 S17462	glutathione transf
44	253	4.7	210	2 S71958	glutathione transf
45	251.5	4.7	208	2 S41933	glutathione transf

ALIGNMENTS

RESULT 1

JC5234

amine oxidase (copper-containing) (EC 1.4.3.6) AOC2 precursor - human

C:Species: Homo sapiens (man)

C:Date: 13-Mar-1997 #sequence-revision 13-Mar-1997 #text-change 18-Feb-2000

C:Accession: JC5234

R:Zhang, X.; McIntire, W.S.

Gene 179, 279-286, 1996

A:Title: Cloning and sequencing of a copper-containing, topa quinone-containing mon

A:Reference number: JC5234; MUID:97128319; PMID:8972912

A:Contents: Placenta

A:Accession: JC5234

A:Molecule type: mRNA

A:Status: preliminary

A:Residues: 1-763 <ZHA>

A:Cross-references: GB:U39447; NID:q1399031; PIDN:AAC50919.1; PID:q1399032

C:Comment: This enzyme catalyzes the oxidation of primary amines to the correspondi

Cu(II) and 1 molecule of covalently-bound topa quinone. It also can scavenge circ

C:Genetics:

A:Gene: GDB:AOC2; DAO2

A:Cross-references: GDB:4562632

A:Map position: 17q21-17q21

C:Superfamily: amiloride-binding protein

C:Keywords: copper; metalloprotein; oxidoreductase; quinoprotein; topaquinone

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-763/Product: amine oxidase (copper-containing) #status predicted <MAT>

F:474/Modified site: copper (His) #status predicted

F:471/Modified site: topaquinone (Tyr) #status predicted

Query Match 74.1% Score 3967; DB 2; Length 763;

Best local similarity 99.9%; Pred. No. 3.5e-293;

Matches 735; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	263	QGGGGGSPQLPHCPSPSPSAQPTWHPGQSLFADLSRELLTAYMRELLORGLGLVDA	322
DB	28	RGGGGGSPQLPHCPSPSPSAQPTWHPGQSLFADLSRELLTAYMRELLORGLGLVDA	87
QY	323	QARSDNCVSEVQLPPKAAALAHLDGSPPARALAIYFGRQDPNSELVGPPLP	382
DB	88	QARSDNCVSEVQLPPKAAALAHLDGSPPARALAIYFGRQDPNSELVGPPLP	147
QY	383	HPSYMRDVTYERRGGLPYHRRPVLFQYLDIDIMIFNRELPOASGLIHCCFYKHGRN	442
DB	148	HPSYMRDVTYERRGGLPYHRRPVLFQYLDIDIMIFNRELPOASGLIHCCFYKHGRN	207
QY	443	LVTMTTPRGLQSDRATWGLYINISAGFLLHVGLELLVNHKALDPAWTOXKYQ	502
DB	208	LVTMTTPRGLQSDRATWGLYINISAGFLLHVGLELLVNHKALDPAWTOXKYQ	267
QY	503	GRYDLSAQLAEAGLVNVVLLIPDNGTGSMSLSKSPVPGAPAPLQFYDQGRFVSOG	562
DB	268	GRYDLSAQLAEAGLVNVVLLIPDNGTGSMSLSKSPVPGAPAPLQFYDQGRFVSOG	327

QY 563 SNVAALMFSEFGLAFSGPRIFDVRFGELVYEISLQELALATIGGNSPAMTTRIVYG 622
| | | | |
Db 328 SRYAASLMFTFSGLGAFSGFRITDFVRFGELVYEISLQELALATIGGNSPAMTTRIVDG 387
| | | | |
QY 623 GFGMGKVTPLTRGVGCPLYATVVDHMFLLSESQAPTIRDAFCVEEQNGFLRLRRHSDL 682
| | | | |
Db 388 GFGMGKVTPLTRGVGCPLYATVVDHMFLLSESQAPTIRDAFCVEEQNGFLRLRRHSDL 447
| | | | |
QY 683 YSHYFGGLAEVLVLRSMSTLLNVDYWDTVFHPGSAIEIRFYATGYISSAFLFGATGKY 742
| | | | |
Db 448 YSHYFGGLAEVLVLRSMSTLLNVDYWDTVFHPGSAIEIRFYATGYISSAFLFGATGKY 507
| | | | |
QY 743 GNQVSHTLTGTVTHSAHFVLDLDAGLENNWMAEDMWVFVMAVWSPEHOLOQLQVTEKK 802
| | | | |
Db 508 GNQVSHTLTGTVTHSAHFVLDLDAGLENNWMAEDMWVFVMAVWSPEHOLOQLQVTEKK 567
| | | | |
QY 803 LLEMEOQAFLVGSAATPRYLTYLASNHSNKMGHPRGYRIOMLSFACEPLPONSSMARGFSN 862
| | | | |
Db 568 LLEMEOQAFLVGSAATPRYLTYLASNHSNKMGHPRGYRIOMLSFACEPLPONSSMARGFSW 627
| | | | |
QY 863 ERYQLAVTORKEEPPSSSVFNQNDPMAPYVDFSDFINNETIAGDLVAMVTAAGFLHIPIH 922
| | | | |
Db 628 ERYQLAVTORKEEPPSSSVFNQNDPMAPYVDFSDFINNETIAGDLVAMVTAAGFLHIPIH 687
| | | | |
QY 923 AEDIPTVVGNVCVGFRLRPNFEDEDPDSFYADSISTYFRGDODACACEVNPLACLPOAAA 982
| | | | |
Db 688 AEDIPTVVGNVCVGFRLRPNFEDEDPDSFYADSISTYFRGDODACACEVNPLACLPOAAA 747
| | | | |
QY 983 CAPDLPAFSGHGFSHN 998
| | | | |
Db 748 CAPDLPAFSGHGFSHN 763
| | | | |

RESULT 2

A54411
amine oxidase (copper-containing) (EC 1.4.3.6), serum, precursor - bovine

C:Species: Bos primigenius taurus (cattle)
C:Date: 09-Sep-1994 #sequence_revision 06-Feb-1995 #text_change 18-Feb-2000
C:Accession: A54411, B38081; A48242; S65408
J:RMU, D.: Mezdlinradsky, K.F.; Adams, G.W.; Mayer, P.; Hines, W.M.; Burlingame, A.L.; Smith, J.P.
J: Biol. Chem. 267, 1979-1982, 1992
A>Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine oxidase.
A:Reference number: A54411; MUID:94193686; PMID:8144587
A:Accession: A54411
A:Molecule type: mRNA }
A:Residues: 1-762 <MUA>
A:Cross-references: GB:S65983; NID:9546215; PIDN:AAB30397.1; PID:9546216
R:MU, D.: Janes, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Kilman, J.P.
J: Biol. Chem. 267, 1979-1982, 1992
A>Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine oxidase.
A:Reference number: A38081; MUID:92235001; PMID:1569055
A:Accession: B38081
A:Molecule type: protein
A:Residues: 463-465, 'D', 467-473, 'X', 475-485 <MU2>
R:Janes, S.M.; Mu, D.; Wenner, D.; Smith, A.J.; Kaur, S.; Malby, D.; Burlingame, A.L.; Science 248, 981-987, 1990
A>Title: A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of copper amine oxidase.
A:Reference number: A48242; MUID:90260646; PMID:2111581
A:Accession: A48242
A>Status: preliminary
A:Molecule type: protein
A:Residues: 468-472 <UAN>
R:de Blaise, D.; Agostinelli, F.; de Matteis, G.; Mondovl, B.; Morpurgo, L. Eur. J. Biochem. 237, 93-99, 1996
A>Title: Half-of-the-sites reactivity of bovine serum amine oxidase. Reactivity and chemotaxis of the enzyme.
A:Reference number: S65408; MUID:96203913; PMID:8620899
A:Accession: S65408
A:Molecule type: protein
A:Residues: 463-469, 'X', 471-487 <DEB>
C:Superfamily: halohalide-binding protein
C:Keywords: copper; glycoprotein; oxidoreductase; quinoloprotein; topaquinoxone
E:1-16/Domains: signal sequence #status predicted <SIO>

F.117-762/Product: amine oxidase (copper-containing) #status predicted <MAT>
F.136,231,665/Binding site: carboxylate (Asn) (covalent) #status predicted
F.470/Modified site: topaquinone (Tyr) #status experimental

Query Match	Score	DB 2;	Length
62.18;	3326;		762;

Matches 606; Conservative 57; Mismatches 70; Indels 8; Gaps 1;

263 QGGDGE-----PSQLPHCPVSQPTWTHPGSQLFADLSRELTAVMRELTORL 314
Y

b 19 EGGVGSEEGVGKQCHPSLP RCPSPSPDQPTWTHPDQSLFADLSREELTTVMSELTQQL 78

315 GPGLVDAQARPSDNCVFSVELQLPRKAAALAHLDRGSPPPAREALAIVFFGRQPQPNVS 374

b 79 GPDLVDAQARPSDNCVFSEVLQLPPKAAALAHLDRGSPPPAREALAIVFEGGPQPNVT 138

375 ELVVGPLPHPSYMRDVTVERHGGPLPYHRRPVLFOEYLDIDOMIFNREL PQASGL LHCC 434

b 139 ELVVGPLPQPSYMRDVTVERHGGPLPYRRPVLREYLDIDQMIENRELPGAAGVLHHC 198

435 FYKHGRNLVTMTAPRGLSGDRATWFGLYYNISGAGFFLHHVGLLELVNHKALDPARW 494

b 199 SYKGGCKLLTMSAPRGVQSGDRSTWFGIYYNITKGGPYLHPVGLELLVDHKALDPADW 258

495 TIQKVFYQGRYYDSLQLEAQFEAGLVNVLIPDNGTGGSMLSKSPVPPGPAPPLQFYPO 554

259 TVQKVFEGGRYYENLAQLLEEQFEAGQVNVVIPPDDGTGGFWSLSKQVPPPGPTPLQEHPO 318

555 GPRFSVQGSRASSLWTFSTFGLGAFSGPRIFDVRFQGERLVEISLQELALTYGNSPAA 614

319 GPRFSVQGNRVASSLWTFSGLGAFSGPRVFDFVRFGGERLAYEISLQEAGAVYGGNTPA 378

615 MTRRYVDGGFGMGKYYTPTLRGVDCPYLATYVDWHFLLSQAPKTIKDAFCVFQONQGLP 674

379 MLTRYMDSGFGMGYFATPLIRGVDCPYLATYMDWHFVESQTPKTLHDAFCVFEEQNKGLP 438

675 LRRHSDLYSHYEGGLAETVLVRSMSTLLNDBYWDTVFHPGAEIRFYATGISSAF /34

439 LRRHSDFLSHYFGGVAQTVLVERSVSTMLNBDYVWDMVFYPNGAIEVKLHATGYISSAF 498

735 LFQATGKRYGNQVSEHTLGTVHTSHAEKVDLDVAGLENNVWAEQMFVPMAPWSPHEQL /94

499 LEGARRRYGNQGEHTLGPRVHTSHAHYKVDLDVGGLNNWVAEDMAFVPIAIPWSPEHQI 558

795 QRLQVTRKLLLEMEEQAAFLVGSATPRILYLASNHSNKGWHPKGRIRIQLMWSFAGEPLQNS 854

339 QRLDVIKQULEIEEQAAPUGGASPKILILASAKUSNAWGHFKGIRKYLVSFAAGFMQNS 010

833 SMARCSWEXIQLAVIQRNEEPSSSVFNQNDPWAPIVDISDIINNEIISGNDLVAMVI 314

019 PMEKAF SWGR I Q L A I I Q R K N E I E P S S S S V E N U N D F M I P I V D F S D F I N N E I I A G N D E V A M V I 0/0

913 AGT LNFHAEIDFNIVVNSGVF LNF INFLDEDFSELSRDSGLFVDDYDNRACLVNFD 274

0 9 AGF LNF FHE DIF N Y I V O N G V O F L N F I N E D Q E F S W S S D S I A F N E O J C R O S C E I N E D / 20

xy	3/3	ACBFGHNNNNNN	WFFNNNNNN	333
				:

133 ACPLFMMALCPLDPLVLSNOOL 133

RESULT 3.

amine oxidase (copper-containing) (EC 1.4.3.6), kidney, precursor - rat

4;Contains: amloride-binding protein, short form

```

-;species: rattus norvegicus (norway rat)
-;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Feb-2000

```

linguegla, E.; Renard, S.; Voilley, N.; Waldmann, R.; Chassande, O.; Lazdun

DOI: 10.1002/1097-4644(200105)210:3<387::AID-JEMB2103873030>3.0.CO;2-1

DATE: 03/07/2008 TIME: 00:00:00

A:Accession: S36847
A:Molecule type: mRNA
A:Residues: 1-746 <LIN2>
A:Cross-references: EMBL:X73911; NID:g395064; PIDN:CAA52116.1; PID:g395065
A:Accession: S36848
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 529-746 <LIN2>
A:Cross-references: EMBL:X73912; NID:g395066; PIDN:CAA52117.1; PID:g395067
R:Verity, K.; Fuller, P.J.
Am. J. Physiol. 266, C1505-C1512, 1994
A:Title: Isolation of a rat amiloride-binding protein cDNA clone: tissue distribution and
A:Reference number: 151904; MUID:94295632; PMID:8023885
A:Accession: 151904
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 530-586, 'T', 588-632 <VER>
A:Cross-references: GB:S70383; NID:9547214; PIDN:AAB31157.1; PID:9547215
A:Note: In Genbank entry S70383, release 116.0, the source is designated as Rattus sp.
C:Superfamily: amiloride-binding protein
C:Keywords: alternative initiators; oxidoreductase
F:1-746/Product: amine oxidase (copper-containing), kidney, precursor #status predicted
F:529-746/Product: amiloride-binding protein short form #status predicted <SHORT>

Query Match 26.1%; Score 1398.5; DB 2; Length 746;
Best Local Similarity 40.3%; Pred. No. 7, 1e-98;
Matches 294; Conservative 127; Mismatches 266; Indels 43; Gaps 13;

QY 291 OSOLFADLSRELTAVMRFLTORLGPGLVDAQAARPSDNCVSEVLQPPKAAALHADR 350
DB 29 KAREFADLSQELKAVHSFLMNEELGLOPSKEPTLAKNSVFLIEMLPKKHKLKLE 88
QY 351 GSPPEAREALAIYFFGROPQPNVSELVYGLPHPSYMRDYVERHGGPLRYHRRPVLEOE 410
DB 89 GRKGPNEARAVIFFGADQDPNVTFAVGLPRPYTRALS-PPRGHLLSWSSPSIAE 147
QY 411 YLIDIDQIFNRELPOASGLAH-----CCFYKRRGNLYTMTARGLSGDRATMG 463
DB 148 Y----DLLYH-TLKRATMPRLHQFLDTGFSFLGCDRCLEFTDVAPRGVASSGORSWEI 202
QY 464 LYVNSAGPFLHNVGELVNHKALDPAWTKQVFOGQYUUSLQKQFAGLVNV 523
DB 203 VQRVVE--GIFLHPTGELILDHGSTVDQMRVQDLMTNGKFTNNPEELAKYAVGEVD 260
QY 524 VLIDP---NGTGSMSLSPY-----PPGAPRLQFYPOGFRFSVGSRAVS 568
DB 261 VLLEDPLPNGE-----KPLFSSYKPRGEFHTLPVNVAGPVGSGRYLLEGWTVL 315
QY 569 LMTFSGLGAFSGPRIDVDFRQGERLYVEISLOALAIYGGNSPAAMTTRIVDGGFGMK 628
DB 316 GMSFSYLRSSSGLOIFENVLGGERVAYEVSQEAVALYGGHTPAGMOKYIDVWGGLGS 375
QY 629 YTTTLRGVDPPIYATVDMHFLLESQAKTIIRDAFCVEQNOGLRLRHSDYS--H 685
DB 376 VTHELAPRIDCPETATFIDAFHYIDSQGVHPRALCLFEMPTGVLRLRHNSNFKGPN 435
QY 686 YFGGLAETVLVVRSMSTLANTDYWDVTFHPSGAIEIFYATGYISSAFLEGATGKYNQ 745
DB 436 FYAGLKVVLVLRITSTYVNDYIMDFIFYSNGVEMAKMAHTGYHAFFYREGLRHGR 495
QY 746 VSEHTLGVTHSAHFKVDLDVAGLENNVMAEDVFMVAVPASEHOLQRLQVTRKLE 805
DB 496 LQTHLGNIRHVLVHYRVMDVAGTKNSFQTLTKMLEMLTNWSPSHSLYQPTLQOTYXS 555
QY 806 MEEQAAFLVGSATPRYLTLASNSKMGHPRGYRIOMLSFAGEPLPONSMSARGFSEWERY 865
DB 556 QEHQAAFFRGOTLPKYLFLFSSPOKCKWGRHRSYRLQIHSMAQVLPQGOEERAVTARY 615
QY 866 QLAATQREKEEPPSSSVENQNDPAAPTVDFSDFI--NNETIAGKDLVAVWTAGFLIIPAE 924
DB 616 PLAVTKYRESERYSSTLVNQNDPMPDPVYVEEFELRNNEIEDDLVAVWTAGFLIIPHE 675
QY 925 DIPNTVTVNGVGFLLRPYNEFDEDBSPYSADSIYFRGDDQDAGACEVNPVLACLPQAA 984

DB 676 DVNNTATPGNSGFLLRPFPEPDESLASRDIVY-WPQDKLNVV--RWLPEDRRCL 732
QY 985 PDLPAFSGG 994
DB 733 VS-PFSSYNG 741

RESULT 4
A54053
amine oxidase (copper-containing) (EC 1.4.3.6) ABP1 precursor - human
N:Alternate names: amiloride-binding protein; diamine oxidase; histaminase
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 18-Feb-2000
C:Accession: A54053; A54410; I38708; A38276; S42495
R:Chassande, O.; Renard, S.; Bardry, P.; Lazdunski, M.
J. Biol. Chem. 269, 14484-14489, 1994
A:Title: The human gene for diamine oxidase, an amiloride binding protein. Molecu
A:Reference number: A54053; MUID:94237856; PMID:8182053
A:Accession: A54053
A:Molecule type: DNA
A:Residues: 1-751 <CHA>
A:Cross-references: EMBL:X78212; NID:9463242; PIDN:CAA5046.1; PID:9463243
R:Novotny, W.F.; Chassande, O.; Baker, M.; Lazdunski, M.; Bardry, P.
J. Biol. Chem. 269, 9921-9925, 1994
A:Title: Diamine oxidase is the amiloride-binding protein and is inhibited by amilo
A:Reference number: A54410; MUID:94193685; PMID:814586
A:Accession: A54410
A:Molecule type: Protein
A:Residues: 20-27, 'A', 29-39 <NOV>
A:Cross-references: PIDN:AAB30395.1; PID:g546213
A:Experimental source: placenta
A:Note: Sequence extracted from NCBI backbone (NCBI:P:146046)
R:Zhang, X.; Kim, J.; McIntire, W.S.
Biochem. Genet. 33, 261-268, 1995
A:Title: cDNA sequences of variant forms of human placenta diamine oxidase.
A:Reference number: I38708; MUID:96113540; PMID:8595053
A:Accession: I38708
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-331, 'F', 333-751 <RES>
A:Cross-references: EMBL:U11862; NID:9535353; PIDN:AAC50270.1; PID:9535356
R:Barbry, P.; Champé, Sci. U.S.A. 87, 7347-7351, 1990
A:Title: Human kidney amiloride-binding protein: cDNA structure and functional exp
A:Reference number: A38276; MUID:91017502; PMID:2217167
A:Accession: A38276
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-265, 'K', 267-793 <RES>
A:Cross-references: GB:M5602; GB:M36335; NID:g87653; PIDN:AAA58358.1; PID:g17796
A:Note: This sequence has been corrected in reference A54053
C:Genetics:
A:Gene: GDB:ABP1
A:Cross-references: GDB:127105; OMIM:104610
A:Map position: 7q31-7q32
A:Introns: 524/1; 619/2; 663/3
C:Function:
A:Description: oxidation of putrescine and histamine
C:Superfamily: amiloride-binding protein
C:Keywords: homodimer; oxidoreductase; quinoprotein; topaquinone
F:1-19/Domains: signal sequence #status predicted <SIG>
F:461/Modified site: topaquinone (Tyr) #status predicted

Query Match 26.0%; Score 1394; DB 2; Length 751;
Best Local Similarity 39.8%; Pred. No. 1, 1e-97;
Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;

QY 285 PWTBQGSOLFADLSRELTAVMRFLTORLGLGLVDAQAARPSDNCVSEVLQPPKAA 344
DB 23 PGLTRKAGVFDLSQELKAVHSFLMSKKELELRPSSSTTAKNTVFLIEMLPKKHYV 82
QY 345 LAHDRGSPPEAREALAIYFFGROPQPNVSELVYGLPHPSYMRDYVERHGGPLRYHRR 404


```

Db      83 LRFIDKGRHVRRAARAVIFFGQDEHPVTEFAVGLPFGPCYMRALS - PRGYGSSMSR 141
QY      405 PVLQOEYLDIDQMIFNRLRLQASGLHH-----CCFYKRGRLVMTTAPRGLQSGD 457
Db      142 PISTAEY-----ALLYHTLOEATKPLHQFLINTGTFSDQCHDRCLAFITDVAIPRGVASQ 196
QY      458 RAEFGLYYNISGAGFLLHGLVLLVNHKALDPAWRTIOKVEYOGGRYDLSAOLAEQF 517
Db      197 RRSLLIIOIRYVE--GYFLHPITGLLELDHGSTDAGHMAVEQWVNGKTYGSEELAKRYA 254
QY      518 AGLVNVLIPDNGTSGMSKSPVP-----GPAP-----PIQYPOGRFYSVOG 562
Db      255 DGEVDVAVVLEPRLPGKGHDSTEEPLFSSSHKPRGDFSPPIHVSQPRLVQHPGRFREG 314
QY      563 SRVASLMTSEFGLGARGSPRIEPRQGEVLVEISQELALITGSGSPRAAMTRRYDG 622
Db      315 NAVLYGKSFELRURSSGLDVLVNHFGGERIAEVSQVQVALYGGHTPMQTKYLDV 374
QY      623 GFGMGKYYTPPLTRGVDCPYLATVYVWHLLESQAPKTRDAFCVFEQNOGQPLRRHSD 682
Db      375 GWSGVSYTHLAPGIDCEPTEFIDTFHYADDPVHPRALCLFEMPTGYPLRRHNSN 434
QY      683 YS---HFFGLAEVLYVRSNSTLNDYVDYVPHPSGATEIRPYATGYISAFELGAT 739
Db      435 FKQGFNFYAGKGVLRITSTVYNDYIMDFIYPNGVMEAKMHATGYHATEFYEPG 494
QY      740 GKYGNQSEHTLGTVHTHSAFKYDLVAGLENVWMAEDMVFVMAVWSPSEHOLQRY 799
Db      495 LRHGTIRLTHLIGNHTLHYRDLDAVAGTKNSFQTLQMLNTNTNWSRHRVQPTL 554
QY      800 TRKLEMEQOAFVLSATPRYLVLASNHSNMKGHPRGYRIQMSFAGEPQNSMAR 859
Db      555 EQTQSMERQOAFRRFKRLPRYLFTSPQENPMGHKRSYRIQHSMAQVLPFGQEOA 614
QY      860 FSWRRYQALAYQKEEPPSSSVNNDPAPYVDSDFI--NETTIAGKDLVAVYTAGFL 918
Db      615 ITWARYPLAVYKRESELCSSSIYHNDPMHPPVYFEOFLNNENIEMEDVAVYTAGFL 674
QY      919 HIPAEDIPNTVYVNGVGFELRPYNFEDEPFSYASDIYFRGDQDAGCEVNP 978
Db      675 HIPSEDIPTNATGNSVGLFRLRFNFPEDPSLASRDTYLV--WRDNGPNTYVO--RWIP 731
QY      979 QAAACAPDLPAFSGHG 994
Db      732 EDRDCSMP--PFSYNG 746

RESULT 5
A26484
glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma japonicum) (fragment)
C:Species: Schistosoma japonicum
C>Date: 25-Oct-1987 #sequence, revision 30-Sep-1989 #text, change 03-Feb-1994
C:Accession: A94139; A26484; A28315
R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 8703-8707, 1986
A:Title: Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 129/J
A:Reference number: A94139; MUID:87041520; PMID:3095841
A:Accession: A94139
A:Molecule type: mRNA
A:Residues: 1219 <SMIT>
R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 6541, 1987
A:Reference number: A94181
A:Contents: annotation; revision to residues 210-219
C:Superfamily: glutathione transferase
C:Keywords: transferase

Query Match      21.3%; Score 1140; DB 2; Length 219;
Best Local Similarity 98.6%; Pred. No. 4, 4e-79;
Matches 213; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      39 PILGKWKIKGLVQPTRLLEYLEEKYEENHYERDEGDKWRNKKFELGLEFNPPLYIDGD 98

```

```

Db      4 PILGKWKIKGLVQPTRLLEYLEEKYEENHYERDEGDKWRNKKFELGLEFNPPLYIDGD 63
QY      99 VKLQSMALIRYADKHNMLGSPKERAELISMLEGAVLDIRGVSRIRYASDFETLKADF 158
Db      64 VKLQSMALIRYADKHNMLGSPKERAELISMLEGAVLDIRGVSRIRYASDFETLKADF 123
QY      159 LSKLPEMLKMEEDRLSKHTYINGDHVTHPPDMLDALVLYLMDPMCLDAPFKVSEFKR 218
Db      124 LSKLPEMLKMEEDRLSKHTYINGDHVTHPPDMLDALVLYLMDPMCLDAPFKVSEFKR 183
QY      219 IEAIPQIDKYLKSSKYIAMPLOGMATFGGDDHPK 254
Db      184 IEAIPQIDKYLKSSKYIAMPLOGMATFGGDDHPK 219

RESULT 6
A45556
glutathione S-transferase - fluke (Schistosoma mansoni)
C:Species: Schistosoma mansoni
C>Date: 22-Apr-1993 #sequence, revision 18-Nov-1994 #text, change 03-May-1996
C:Accession: A45556
R:Wright, M.D.; Harrison, R.A.; Melder, A.M.; Newport, G.R.; Mitchell, G.F.
Mol. Biochem. Parasitol. 49, 177-179, 1991
A:Title: Another 26-kilodalton glutathione S-transferase of Schistosoma mansoni.
A:Reference number: A45556; MUID:92131046; PMID:1775156
A:Accession: A45556
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-218 <WRI>
A:Note: sequence extracted from NCBI backbone (NCBI:77814)
C:Superfamily: glutathione transferase

Query Match      18.4%; Score 985; DB 2; Length 218;
Best Local Similarity 82.9%; Pred. No. 2, 7e-67;
Matches 180; Conservative 19; Mismatches 18; Indels 0; Gaps 0;

QY      38 SPILGKWKIKGLVQPTRLLEYLEEKYEENHYERDEGDKWRNKKFELGLEFNPPLYIDG 97
Db      2 APRKQYWKIKGLVQPTRLLEYLEEKYEENHYERDEGDKWRNKKFELGLEFNPPLYIDG 61
QY      98 DVKLTQSMALIRYADKHNMLGSPKERAELISMLEGAVLDIRGVSRIRYASDFETLKADF 157
Db      62 DVKLTQSMALIRYADKHNMLGSPKERAELISMLEGAVLDIRGVSRIRYASDFETLKADF 121
QY      158 FLKSLPEMLKMEEDRLSKHTYINGDHVTHPPDMLDALVLYLMDPMCLDAPFKVSEFKR 217
Db      122 FLNQLPGMLKMEEDRLSKHTYINGDHVTHPPDMLDALVLYLMDPMCLDAPFKVSEFKR 181
QY      218 IEAIPQIDKYLKSSKYIAMPLOGMATFGGDDHPK 254
Db      182 IEAIPQIDKYLKSSKYIAMPLOGMATFGGDDHPK 218

RESULT 7
A45523
glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma mansoni)
C:Species: Schistosoma mansoni
C>Date: 03-Jun-1993 #sequence, revision 03-Jun-1993 #text, change 18-Jun-1999
C:Accession: A45523; A44941
R:Trottein, F.; Kieny, M.P.; Verwaerde, C.; Torpier, G.; Pierce, R.J.; Balloul, J.M.;
Mol. Biochem. Parasitol. 41, 35-44, 1990
A:Title: Molecular cloning and tissue distribution of a 26-kilodalton Schistosoma man
A:Reference number: A45523; MUID:90348710; PMID:2365266
A:Accession: A45523
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <TRO>
A:Cross-references: GB:J31106; NID:q161004; PIDN:AAA29888.1; PID:g552241
R:Henkle, K.J.; Davern, K.M.; Wright, M.D.; Ramos, A.J.; Mitchell, G.F.
Mol. Biochem. Parasitol. 40, 23-34, 1990
A:Title: Comparison of the cloned genes of the 26- and 28-kilodalton glutathione S-tr
A:Reference number: A44941; MUID:90271935; PMID:1693415

```

A:Accession: A44941
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 8-218 <HNN>
A:CROSS-references: GB:M26914, GB:M26913, NID:g161006, PIDN:AAA29689.1, PID:g1610075
A:Superfamily: glutathione transferase
C:Keywords: transferase

Query Match	17.2%;	Score 922;	DB 2;	Length 218;
Best Local Similarity	80.2%;	Pred. No. 1.6e-62;		
Matches 174;	Conservative 15;	Mismatches 28;	Indels 0;	Gaps 0

```
QY      38 SPILGWKXKGLVOPRLLLEKEEYEEHLYEREGGDKWRKKFELGLEFPNNPYIDG 90
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2 APRFGYWKXVGLVPRLRLLEHEETYEERAYDRNEIDAWSNDKFELGLEFPNPIYIDG 61
```

Dy 98 DYKLTQSMAIIRYIADKHNMGGSPKERAEISMLEGAVLDIRYGVSRIASKDFETIKVD 157
| | | | | | | | | | | | | | | | | | | | | : : : : :
Db 62 DFKLTQSMAIRYIADKHNMGGACPKERAISMLEGAVLDIRMGVLRIAINKEYETIKVD 121

Oy 158 FLSLPRLKMFEDRSHKTILNGDH7HPDFMLYDALDVLVLYMDMCLDAEPKLVSFKK 21
||:||| |||||||||::|||::| | | | | | | | | | | | : : |||
Db 122 FLLPLGRLEKFEDRLSNKITLNGNCYTHPDFMLYDALDVLVLYMDSQLNEPKLVSEKK 181

QY 218 RIEALPQIDKYLKSSKYLAMPLOGQATFGGSHPPK 254
 || : ||| || ||: || ||||| ||||| |||
 Db 182 CIEDLPQIKNYLNSSRYIWPLOGWDATFGGCDTPK 218

RESULT 8
A48388
Glutathione S-transferase - liver fluke (fragment)

C.Species: fasciola hepatica (liver fluke)
C.Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
C.Accession: A48388
R.Muro, A. Rodriguez-Medina, J.R. Hillyer, G.V.

A: Accession: A48388
M: Reference number: A48388; MUID:9328188; PMID:7682383
A: Title: Sequence analysis of a fascicula hepatica glutathione S-transferase cDNA clone
M: J. Trop. Med. Hyg. 46, 437-460, 1993
A: Accession: A48388

A;status: preliminary
A;molecule type: nucleic acid
A;Residues: 1-209 <MUR>
A;Experimental source: Puerto

Query Match 12.4% Score 663 DB 2: length 209;
A:Note: sequence extracted from NCBI backbone (NCBIN:129136, NCBI:129135)
C:Superfamily: glutathione transferase

[illegible]

Db 1 IRGLOQPVRLLEYLEGEYEENHLYGRDDEKMFQDFKFMGDLPLNLPYIIDCKKLTQSV 60

Db 61 AIMRYIADKHGMIGSTPEEARARISMIEGAMDLRNGFVRQCYNPKKEEVKGGDYLLKELPTT 120

D6 121 LKMSDELGRHYLTGSTVSHVDPMVEALDCIRLAPQCLEDFPKLKEFKSRIEDLPKI 18

OY 226 DKYLKSSKYIAMPLOGMOATFGGDDHP 253

Db 181 KAYMESEKFIKWP LNSWIASFGGDAAP 208

RESULT 5
S33860
glutathione transferase (EC 2.5.1.18) mul - mouse
N: Alternate names: glutathione transferase major chain; glutathione transferase mu-8.7

C/Species: Mus musculus (house mouse)

C:Date: 08-Dec-1993 #sequence_revision 07-Oct-1994 #text_Change 18-Jun-1999
C:Accession: S33860; A34159; A28946; A20831; I24735; S16933; S7312
R:Reinhart, J.; Pearson, W.R.
Arch. Biochem. Biophys. 303, 383-393, 1993
Article: The structure of two murine class-mu glutathione transferase genes coordin
A:Reference number: S33860; MID:93290350; PMID:8512323

N:Molecule type: DNA
A:Residues: 1-218 <RET>
A:Cross-references: GB:LI3448

R. Townsend, A.J.; Goldsmith, M.E.; Pickett, C.B.; Cowan, K.H. *J. Biol. Chem.* 264, 21582-21590, 1989

A. J. Title: Isolation, characterization, and expression in *Escherichia coli* of two mur

N: Accession: A34159
A: Molecule type: mRNA
A: Residues: 1-218 <TOM>

R; Pearson, W.R.; Reinhardt, J.; Slisk, S.C.; Anderson, K.S.; Adler, P.N. J. Biol. Chem. 263, 13324-13332, 1988

A; Title: Tissue-Specific Induction of murine glutathione transferase mRNAs by butyl

A:Contents: clone pGT 875
A:Accession: A28946
A:Molecule type: mRNA

A: Cross-references: GB:J039352; NID:q193687; PIDN:AAA37747.1; PID:q309278
R: Pearson, W.R., Wandle, J.J.; Morrow, J.F.; Benson, A.M.; Talalay, P.
J. Biol. Chem. 256, 2052-2062, 1983

A:Reference number: A92411; MUID:83109018; PMID:6832548
A:Accession: A20831
A:Molecule type: Protein

R:Mannerlyk, S.; Alin, P.; Gutherberg, C.; Jonsson, H.; Tahiri, M.K.; Warden, M.; Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
Title: Identification of three classes of cytosolic glutathione transferase common

A:Accession: I24735
A:Molecule type: protein
A:Residues: 2-7, 'X', 8-13, 'X', 15-23, 'X', 25 <MAN>
Reviewed by A. D. Meteljan I. T

Biochem. J. 277, 501-512, 1991
A: Title: Hepatic glutathione S-transferases in mice fed on a diet containing the adjuvant elution of the glutathione-Sepharose affinity matrix.
106072

A:Accession:SI16933
A:Molecule type: protein
A:Residues: 2-36 <HA1>

Arch. Biochem. Biophys. 331, 104-116, 1996
A:Title: Elevations of hepatic quinnone reductase, glutathione, and alpha- and mu-c
A:Reference number: S71311; MUID:96268466; PMID:8660689

A: Molecule type: protein
A: Residues: 97-108 <FER>
C: Genetics:

	status	experimental	<MAT>
A:introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3			
C:Superfamily: glutathione transferase			
C:Keywords: dimer; transferase			

Query March	9.18	Score 489.5	DB 2	Length 218
Best Local Similarity	44.48	Pred. No. 1.3e-29		
Best Overall Similarity	40.0	Winnings 70		
Number of Trials	5			
Gaps	1			

[illegible]

4 ILGMYNVRGLHPHIRMILEXYTDSSYDEKRTYMGAPDFKRSQWMEAFKGLGDFPRLFL 103
95 IDGVKLTQSMALIRIADKHNMCGSPKERAETSMLEGAVLDIRYGVSRIAYSKDEPTL 154

Db 64 IDGSHKRTQSNALIRYLARKHHLDETEERIRADIVENQVMDFPMQIMLCYNDFEKO 123
 QY 155 KVDLSKLPKEMLKMFEDRLSHKTYLNGDHVTHPPDMYDALDVLYMDPKCLDAFPKLYS 214
 Db 124 KPEFLKTIPEKMKLYSEFLGRPFAGDKYTVYDFLAYDILDQYHMEFPKCLDAFPNLRD 183
 QY 215 FKRRIEAIPQIDKYLKSSKYIAMPUL 239
 Db 184 FLARFEGILKRTISAYMKSSRYLTPI 208

RESULT 10

A29794
 N:Glutathione transferase (EC 2.5.1.18) class mu chain 3 - rat
 N:Alternate names: glutathione S-transferase Yb1; ligandin
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 28-Aug-1989 #sequence, revision 28-Aug-1989 #text, change 18-Jun-1999
 C:Accession: A29794; A25510; A24085; B61004; F24735; A33397; S27111; S17167; A26187
 R:Chang, C.; Saltzman, A.G.; Sorensen, N.S.; Hlilpakka, R.A.; Liao, S.
 J. Biol. Chem. 262, 11901-11903, 1987
 A:Title: Identification of glutathione S-transferase Yb-1 mRNA as the androgen-repressed
 A:Reference number: A29794; MUID:87308179; PMID:3040722
 A:Accession: A29794
 A:Molecule type: mRNA
 A:Residues: 1-218 <CHN>
 A:Cross-references: GB:J02810; NID:9204514; PIDN:AAA41293.1; PID:9204515
 R:Lai, H.C.J.; Grove, G.; Tu, C.P.D
 Nucleic Acids Res. 14, 6101-6114, 1986
 A:Title: Cloning and sequence analysis of a cDNA for a rat liver glutathione S-transferase
 A:Reference number: A25510; MUID:86312882; PMID:2875437
 A:Accession: A25510
 A:Molecule type: mRNA
 A:Residues: 1-218 <LAT>
 A:Cross-references: GB:X04229; NID:956337; PIDN:CAA27811.1; PID:956338
 R: Ding, G.U.F.; Lu, A.Y.H.; Pickett, C.B.
 J. Biol. Chem. 260, 13268-13271, 1985
 A:Reference number: A24085; MUID:86033768; PMID:3840477
 A:Accession: A24085
 A:Molecule type: mRNA
 A:Residues: 1-198, 'NC', 201-218 <DIN>
 A:Cross-references: GB:M11719; NID:9204502; PIDN:AAA41287.1; PID:9204503
 A:Experimental source: Clone pGTA/C44
 R:Chang, L.H.; Hsieh, J.C.; Chen, W.L.; Tam, M.F.
 Electrophoresis 11, 589-593, 1990
 A:Title: Identification of rat liver glutathione S-transferase Yb subunits by partial N-
 local isoelectric focusing gel
 A:Reference number: A61004; MUID:91031411; PMID:2226415
 A:Accession: B61004
 A:Molecule type: protein
 A:Residues: 2-26 <CH2>
 R:Mennervik, B.; Altn, P.; Guttenberg, C.; Jansson, H.; Tahr, M.K.; Warholm, M.; Jorndal
 Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
 A:Title: Identification of three classes of cytosolic glutathione transferase common to
 A:Reference number: A24735; MUID:86042634; PMID:3864155
 A:Accession: F24735
 A:Molecule type: protein
 A:Residues: 2-20 <MAN>
 R:Hsieh, J.C.; Liu, L.F.; Chen, W.L.; Tam, M.F.
 Biochem. Biophys. Res. Commun. 162, 1147-1154, 1989
 A:Title: Expression of Yb-1 glutathione S-transferase using a baculovirus expression sys
 A:Reference number: A33397; MUID:89350924; PMID:2669745
 A:Accession: A33397
 A:Molecule type: protein
 A:Residues: 2-21, 212-218 <HS1>
 R:Katusz, R.M.; Bono, B.; Colman, R.F.
 Arch. Biochem. Biophys. 298, 667-677, 1992
 A:Title: Identification of Tyr(115) labeled by S-(4-bromo-2,3-dioxobutyl)glutathione in
 A:Accession: S27111
 A:Molecule type: protein
 A:Residues: 2-20, 83-86, 'X', 88-96, 109-115, 'X', 117-122 <KAT>
 R:Hsieh, J.C.; Huang, S.C.; Chen, W.L.; Lai, Y.C.; Tam, M.F.

Biochem. J. 278, 293-297, 1991
 A:Title: Cysteine-86 is not needed for the enzymic activity of glutathione S-transferase
 A:Reference number: S17167; MUID:91354218; PMID:188338
 A:Accession: S17167
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 2-218 <HS2>
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 9.18; Score 486.5; DB 2; Length 218;
 Best Local Similarity 44.48; Pred. No. 2.2e-29;
 Matches 91; Conservative 38; Mismatches 70; Indels 5; Gaps 1;

QY 40 ILGWYKIKGLVOPRLLEYLEEYERDEG----DKWNRKFELEPPNLPY 94
 Db 4 ILGWYVNRGLTHPIRLLEYSYEEKRYAMGAPDYDRSQMLNEKFKGLDEPNLPYL 63
 QY 95 IDGDKVLTQSNALIRYLADKRNMLGSGPKERAELSMLEGAVLDIRGVSRAYSKEDETL 154
 Db 64 IDGSRKTIQSNALIRYLARKHHLGTEEERIRADIVENQVMDFPMQIMLCYNDFEKO 123
 QY 155 KVDLSKLPKEMLKMFEDRLSHKTYLNGDHVTHPPDMYDALDVLYMDPKCLDAFPKLYS 214
 Db 124 KPEFLKTIPEKMKLYSEFLGRPFAGDKYTVYDFLAYDILDQYHMEFPKCLDAFPNLRD 183
 QY 215 FKRRIEAIPQIDKYLKSSKYIAMPUL 239
 Db 184 FLARFEGILKRTISAYMKSSRYLTPI 208

RESULT 11

S32425
 N:Glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 2) - human
 N:Alternate names: glutathione transferase mu4
 C:Species: Homo sapiens (man)
 C:Date: 07-Apr-1994 #sequence, revision 07-Apr-1994 #text, change 18-Jun-1999
 C:Accession: S32425; S29337
 R:Zhong, S.; Spurr, N.K.; Hayes, J.D.; Wolf, C.R.
 Biochem. J. 291, 41-50, 1993
 A:Title: Deduced amino acid sequence, gene structure and chromosomal location of a no
 A:Reference number: S32424; MUID:93228631; PMID:8471052
 A:Accession: S32425
 A:Molecule type: DNA
 A:Residues: 1-218 <ZHO>
 A:Cross-references: EMBL:X68677; NID:931934; PIDN:CAA48637.1; PID:931935
 A:Note: The authors translated the codon AAG for residue 182 as Arg and CCA for resid
 C:Genetics:
 A:Gene: GSTM4
 A:Cross-references: GDB:134191; OMIM:138333
 A:Map position: 1p13.3-1p13.3
 A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 9.18; Score 486.5; DB 2; Length 218;
 Best Local Similarity 44.18; Pred. No. 2.2e-29;
 Matches 94; Conservative 38; Mismatches 76; Indels 5; Gaps 1;

QY 40 ILGWYKIKGLVOPRLLEYLEEYERDEG----EGDKWRKFELEPPNLPY 94
 Db 4 ILGWYDNRGLAHARLLETTDSYEKKTKTMGAPDYDRSQMLNEKFKGLDEPNLPYL 63
 QY 95 IDGDKVLTQSNALIRYLADKRNMLGSGPKERAELSMLEGAVLDIRGVSRAYSKEDETL 154
 Db 64 IDGAKRTQSNALIRYLARKHHLGTEEERIRADIVENQVMDFPMQIMLCYNDFEKO 123
 QY 155 KVDLSKLPKEMLKMFEDRLSHKTYLNGDHVTHPPDMYDALDVLYMDPKCLDAFPKLYS 214
 Db 124 KPEFLKTIPEKMKLYSEFLGRPFAGDKYTVYDFLAYDILDQYHMEFPKCLDAFPNLRD 183
 QY 215 FKRRIEAIPQIDKYLKSSKYIAMPULGWCATFG 247

A:Residues: 1-218 <CO2>
 A:Cross-references: GB:M66234; NID:9306818; PIDN:AAA57347.1; PID:9306819
 R:Experimental source: HeLa cells
 R:Ross, V.L.; Boad, P.G.
 Biochem. J. 294, 373-380, 1993
 A:Title: Molecular cloning and heterologous expression of an alternatively spliced human
 A:Reference number: S36782; MUID:93384505; PMID:8373352
 A:Accession: S36782
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <ROS>
 A:Cross-references: EMBL:M99422
 R:Experimental source: testis
 R:Comstock, K.E.; Widensten, M.; Hao, X.Y.; Henner, W.D.; Mannervik, B.
 Arch. Biochem. Biophys. 311, 487-495, 1994
 A:Title: A comparison of the enzymatic and physicochemical properties of human glutathione
 A:Reference number: S45684; MUID:94263230; PMID:8203914
 A:Contents: annotation
 R:Taylor, J.B.; Oliver, J.; Sherrington, R.; Pemble, S.E.
 Biochem. J. 274, 587-593, 1991
 A:Title: Structure of human glutathione S-transferase class Mu genes.
 A:Reference number: I37438; MUID:91174774; PMID:2006920
 A:Accession: I37438
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 39-120 <RES>
 A:Cross-references: EMBL:X56837; NID:931936; PIDN:CAA40167.1; PID:931937
 C:Genetics:
 A:Gene: GDB:GSTM4
 A:Cross-references: GDB:134191; OMIM:138333
 A:Map position: 1p13.3-1p13.3
 A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: alternative splicing; dimer; transferase

Query Match 9.0%; Score 483.5; DB 2; Length 218;
 Best Local Similarity 43.7%; Pred. No. 3.7e-29;
 Matches 94; Conservative 38; Mismatches 78; Indels 5; Gaps 1;

QY 38 SPILGYMKIKGLVOPTRILLEEYEEHLYERDEG-----DKWRNKKFELGLEFPNLP 92
 DB 2 SMTLGWMDIRGLAHARILLLEYTDSYEKKRYTMDADPYDRSQWLNKFKLGLDFPNLP 61

QY 93 YIIGDVNLTQSMALIRIADKHNHNLGSGPKERAEISMEGAVDIRGVSRIVASKEPPE 152
 DB 62 YLIDGAHKITQSNALICTIYARKHNLGCTEETKEIRVDLENQAMDVSNQARVCSPDPE 121

QY 153 TLKVDFLSKLPEDMLKMFEDRLSHKTYLNGDVTHTPDVLYDALDVLVMDPWCIDAFPKL 212
 DB 122 KIKPEYLEELPTMQHFQFLGKRPWFVGDKITFVDFLAYVDLHRIEFPNCLDAFPNL 181

QY 213 VSFKKRIFAIPOIDIKYKSSKRYIAMPLOGMQATFG 247
 DB 182 KDFISREGLEKISAYMKSSRFLPKPLYTRVAVMG 216

RESULT 15
 JX0095
 glutathione transferase (EC 2.5.1.18) b - guinea pig
 N:Alternate names: glutathione S-alkyltransferase; glutathione S-aryltransferase; glutathione
 C:Species: Cavia porcellus (guinea pig)
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 18-Jun-1993
 C:Accession: JX0095
 R:Kamel, K.; Oshino, R.; Hara, S.
 J. Biochem. 107, 111-117, 1990
 A:Title: Amino acid sequence of glutathione S-transferase b from guinea pig liver.
 A:Reference number: JX0095; MUID:90236961; PMID:2332413
 A:Accession: JX0095
 A:Molecule type: protein
 A:Residues: 1-217 <KAM>
 C:Comment: Glutathione transferases are a family of enzymes that are multifunctional in
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 9.0%; Score 481.5; DB 2; Length 217;
 Best Local Similarity 45.1%; Pred. No. 5.2e-29;
 Matches 92; Conservative 35; Mismatches 72; Indels 5; Gaps 1;

QY 41 LQYWKIKGLVOPTRILLEEYEEHLYERDEG-----DKWRNKKFELGLEFPNLPYI 95
 DB 4 LGYWNIRGLTHPIRLILEYTNNGYEKKRYNMGADPYDRSQWLNKFKLGLDFPNLPYI 63

QY 96 DGDVLTQSMALIRIADKHNHNLGSGPKERAEISMEGAVDIRGVSRIVASKEPPE 155
 DB 64 DQTHKLTQSNALIRIYARKHNLGCTEETKEIRVDLENQAMDVSNQARVCSPDPEQK 123

QY 156 VDFLSKLPEDMLKMFEDRLSHKTYLNGDVTHTPDVLYDALDVLVMDPWCIDAFPKL 215
 DB 124 AEFLEGIPDKMLFQFLGKLPWFGNKLTYVDFLAYVDLHRIEFPNCLDAFPNL 183

QY 216 KRIRAIPOIDIKYKSSKRYIAMP 239
 DB 184 ISFEGLEKISSYMKSSRFLPKPL 207

Search completed: May 20, 2003, 11:39:29
 Job time : 42.5066 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 23.557 Seconds
(without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357
Sequence: 1 MDWLRLFLFAAAASINNA.....QAACAPDLPFASHGGSFN 998

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3967	74.1	763	1	AOC3_HUMAN
2	3365	62.8	765	1	AOC3_MOUSE
3	3338	62.3	762	1	AOCX_BOVIN
4	3336	62.1	762	1	AOCX_BOVIN
5	2556	47.7	756	1	AOC2_HUMAN
6	1398.5	26.1	746	1	ABP_RAT
7	1394	26.0	751	1	ABP_HUMAN
8	1315	24.5	321	1	AOC3_RAT
9	1144	21.4	218	1	GT26_SCHJA
10	985	18.4	218	1	GT27_SCHMA
11	922	17.2	218	1	GT26_SCHMA
12	719.5	13.4	220	1	GT29_FASHE
13	690	13.2	217	1	GT27_FASHE
14	690	12.9	217	1	GT28_FASHE
15	689	12.9	217	1	GT26_FASHE
16	489.5	9.1	217	1	GTM1_MOUSE
17	489.5	9.1	217	1	GTM1_RAT
18	485.5	9.1	217	1	GTW0_RABIT
19	485.5	9.1	218	1	GTM6_MOUSE
20	484.5	9.0	217	1	GTM2_HUMAN
21	483.5	9.0	218	1	GTM4_HUMAN
22	481.5	9.0	217	1	GTW0_CAVPO
23	476.5	8.9	217	1	GTW0_CRITLO
24	476.5	8.9	217	1	GTW0_MESAU
25	473.5	8.9	217	1	GTW2_MOUSE
26	473.5	8.8	217	1	GTM1_HUMAN
27	471.5	8.8	217	1	GTW3_HUMAN
28	469.5	8.8	217	1	GTW3_RAT
29	463.5	8.7	217	1	GTW2_RAT
30	460.5	8.6	217	1	GTW3_MOUSE
31	455.5	8.5	224	1	GTW3_HUMAN
32	443.5	8.3	224	1	GTW5_MOUSE
33	440.5	8.2	219	1	GTW2_CHICK

34	413	7.7	755	1	AMO_KIEAE
35	397.5	7.4	757	1	AMO_ECOLI
36	395.5	7.4	219	1	GTW1_DERPT
37	377	7.0	683	1	AMOH_ARTGO
38	363	6.8	648	1	AMOI_ARTGO
39	363	6.8	648	1	AMOI_ARTS1
40	340	6.3	638	1	PAOX_ARTGO
41	320	6.0	674	1	AMO_PEA
42	311.5	5.8	666	1	AMO_LENCU
43	310	5.8	671	1	AMO_ASPNG
44	295.	5.5	692	1	AMO_PICAN
45	256.5	4.8	208	1	GTP_DIRIM

ALIGNMENTS

RESULT 1

AOC3_HUMAN STANDARD; PRT; 763 AA.

AC Q16853;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1) (HPAO).
 GN AOC3 OR VAP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=97128319; PubMed=8972912;
 RA Zhang X., McIntire W.S.;
 RT Cloning and sequencing of a copper-containing, topaquinone-
 RT containing monamine oxidase from human placenta.";
 RL Gene 179:279-286(1996).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Lung;
 RX MEDLINE=98317014; PubMed=9653080;
 RA Smith D.J., Salmi M., Bono P., Hellman J., Lau T., Jalkanen S.;
 RT Cloning of vascular adhesion protein 1 reveals a novel
 RT multifunctional adhesion molecule.";
 RL J. Exp. Med. 188:17-27(1998).
 CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE
 CC RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO
 CC PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
 CC INDEPENDENT FASHION. HAS A MONAMINE OXIDASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: MOST STRONGLY EXPRESSED ON THE HIGH
 CC ENDOTHELIAL VENULES OF PERIPHERAL LYMPH NODES AND ON HEPATIC
 CC ENDOTHELIA.
 CC -1- PTM: Topaquinone (TPQ) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue (by similarity).
 CC -1- PTM: N- AND O-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U039447; AAC50919.1; -.

FT	TRANSMEM	7	27	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT	DOMAIN	28	765	EXTRACELLULAR (POTENTIAL).
FT	MOD_RES	471	471	TORAOQUINONE (BY SIMILARITY).
FT	METAL	520	520	COPPER (POTENTIAL).
FT	METAL	522	522	COPPER (POTENTIAL).
FT	METAL	684	684	COPPER (POTENTIAL).
FT	BINDING	673	673	AMILORIDE (BY SIMILARITY).
FT	CARBOHYD	137	137	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	232	232	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	592	592	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	659	659	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	765 AA;	84533 MM;	7489DE67D3DBB44D CRC64;

Query Match	62.8%;	Score 3365;	DB 1;	Length 765;
Best Local Similarity	83.0%;	Pred. No. 4.4e-244;		
Matches 610; Conservative	62;	Mismatches 63;	Indels 0;	Gaps 0;

QY	263	QGGGGGEESSOLPHICSPSPSAOPMTNHPQSOJALSTRELLAVNRFLTORGGPGLVDA	322
Db	28	RSGGGGGLSOLPHICSPSVLPVQPRTHNPSOSQFADLSPEELTAVMSFLTKHJCPGLVDA	87
QY	323	QARPSDNCVFSEVLOLPRKAAALAHLDGSPRPAREALATVFFGSGPOPNVSELYVGP	382
Db	88	QARPSDNCVFSEVLOLPRKAAALAHLDGSPRPAREALATVFFGSGPOPNVSELYVGP	147
QY	383	HPSTWRDVTVERHGGPRLPYHRRPVLFOEYLDIDQIMFNRELPOASGLHNCCFYKRG	442
Db	148	HPSTWRDVTVERHGGPRLPYHRRPVLFOEYLDIDQIMFNRELPOASGLHNCCFYKRG	207
QY	443	LYTMHTARGLSOGGRAPMFGLYNIMISAGFELLYVSGEFLYLNHKLADPARKTQKVE	502
Db	208	LYTMHTARGLSOGGRAPMFGLYNIMISAGFELLYNHLDELIDHKLADPARKTQKVE	267
QY	503	GRYVDSLQLEAQEFAGLVNVLIDPNDGSGSMSKSPRPAPRPLQFPOGPRFSYVG	562
Db	268	GRYVDSLQLEQDFAGLVNVLVNNNGSGSMSKSPRPAPRPLQFPOGPRFSYVG	327
QY	563	SRVASSLTFSFGIAGFSGPRIFDVRFOGERLYVEISLQALATYGGNSPAAMTKRYDG	622
Db	328	SOVSSLSAFSGIAGFSGPRIFDVRFOGERLYVEISLQALATYGGNSPAAMTKRYDG	387
QY	623	GFGMKKTTPLTRGDCRYLATYVVMHFLBESQAKTINDACVPEQONGRLPLRHNSD	682
Db	388	SFGICKYSTPLTRGVDCRYLATYVVMHFLBESQAKTINDACVPEQONGRLPLRHNSD	447
QY	683	YSHYGGGLAEFLVYVRSMSLTINDYVMDTVPHPSGAIREFRYANGYISSAFLEGATGY	742
Db	448	YSHYGGVGVYLVYVRSYSTLNTYDINDVMYHPNGAITEVKNHAICYISSAFFGGEKF	507
QY	743	GNQVSEHTLGVHTHSAHFKYDLDVAGLENNVMAEDMVEPMAVWPSEHOLORLOVTK	802
Db	508	GNRVCAHTLGVHTHSAHFKYDLDVAAGLKNMAAEDMAVPTVPMQREYQORQLOVTK	567
QY	803	LLEMEQAAFLVGSATPRLYLATASHNSKNKGPRGRYIMLSFAGPRLPQNSSMANGBN	862
Db	568	LLETEEAFAFLPGGATPRLYLATASHNSKNMGRRGRYIOLISFAGPRLPOESPIEKAFM	627
QY	863	ERYOLAVTORKEEBSSESVFNQNDPMATVYFSDPINNETJAGKOLVAMVTAAGFLHIP	922
Db	628	GRYHLAVTORKEEBSSESVFNQNDPMATVYFTDPIINNETJAGEDLVAMVTAAGFLHIP	687
QY	923	AEDINTYTVNGVGFELRLPYNFEDBPSFYSAISYIFRGDODAGACEVNPPLACIQA	982
Db	688	AEDINTYTVAGNSVGFELRLPYNFEDBPSFHSADSYIFREGODATACEVNPPLACISQ	747
QY	983	CAPDLPAFSGGFSH 997	
Db	748	CAPDLPAFSGGFSH 762	

ID	ACCY_BOVIN	STANDARD;	PERT;	762 AA.
AC	06406;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Copper amine oxidase, lung isozyme precursor (EC 1.4.3.6) (Amine oxidase [copper-containing]) (BOLA0).			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.			
OX	NCBI_Taxid=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lung;			
RX	MDLID=98151264; PubMed=9492300;			
RA	Hogdall E.V.S., Houen G., Borre M., Bundgaard J.R., Larsson L.-I., Vust J.;			
RT	"structure and tissue-specific expression of genes encoding bovine copper amine oxidases."			
RL	Eur. J. Biochem. 231:320-328(1996).			
CC	-1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) + H(2)O(2).			
CC	-1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit (By similarity).			
CC	-1- ENZYME REGULATION: INHIBITED BY AMIORIDE IN A COMPETITIVE MANNER (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Extracellular (By similarity).			
CC	-1- TISSUE SPECIFICITY: LUNG, SPLEEN, HEART, KIDNEY.			
CC	-1- PFM: Topaquinine (TPO) is generated by copper-dependent autoxidation of a specific tyrosyl residue (By similarity).			
CC	-1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce or send an email to license@isb.slb.ch).			
CC	-----			
DR	EMBL: Y15774; CAA5776.1; -.			
DR	InterPro: IPR000269; CuNH_oxidase.			
DR	Pfam: PF01179; Cu_amine_oxid; 1.			
DR	Pfam: PF02727; Cu_amine_oxidn; 1.			
DR	Pfam: PF02728; Cu_amine_oxidn; 1.			
DR	PRINTS: PR00766; CUDAOXIDASE.			
DR	PROSITE: PS01164; COPPER AMINE OXID 1; 1.			
DR	PROSITE: PS01165; COPPER AMINE OXID 2; 1.			
KW	Oxidoreductase; Copper; TPO; Glycoprotein; Signal.			
FT	SIGNAL	1	16	POTENTIAL.
FT	CHAIN	17	762	COPPER AMINE OXIDASE, LUNG ISOZYME.
FT	MOD_RES	470	470	TOPAQUINONE (By SIMILARITY).
FT	METAL	519	519	COPPER (POTENTIAL).
FT	METAL	521	521	COPPER (POTENTIAL).
FT	METAL	683	683	COPPER (POTENTIAL).
FT	BINDING	672	672	AMIORIDE (By SIMILARITY).
FT	CARBOHYD	136	136	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	231	231	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	293	293	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	617	617	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	665	665	N-LINKED (GLCNAc. . .) (POTENTIAL).
QO	SEQUENCE	762 AA;	84883 MW;	B43D04776744AF2 CRC64;

Query Match	62.3%;	Score 3338;	DB 1;	Length 762;
Best Local Similarity	83.1%;	Pred. No. 4.7e-242;		
Matches 616;	Conservative 42;	Mismatches 75;	Indels 8;	Gaps 1;

QY 263 GGGDGE-----PGLPHCPVSPSAQPWHPGSGOLFADLSREETLAVKREFLTOL 314
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19 EGVGSGEEGVGKOCHSPRPCCSRSPSDQPMWHPDOSQLFADLSREETLAVNSFLTKL 78

Db	319	GPFRSVQGNRRVASSLWTFEFSFGDLGAFSGSPRPVEDVFRQGEERLAEISLQEGAVYGGNTPAA	378
Qy	615	MTTRVYDGGFGMGKYYTPPLTRGVDCPLIATYVDMHFLLESQAPKTIINDAFCYEQONGLP	674
Db	379	MLTRYMDSGFEMGFAPPLRGVDCPLATATYMDHMFVESQTPKTLIDACVFEQNGLP	438
Qy	675	LRRHSDLSYHFGCLAEVTLVWVSMSTLLNYDVMDPTVFHPSCAIEIRFATYISSAF	734
Db	439	LRRHSDLSYHFGCVGQATVLYEFSVSTMLNLYDVMDVTFPNAIEVKLHAICYISSAF	498
Qy	735	LEGATGKGYGNVSEHTLGTVHTSHAFKVDLDVAGLENNWVAEDMVFVPAVWPSPHQ	794
Db	499	LEGAAARRGQNGVGHETLGPVHTSHAHKVDLDVAGLENNWVAEDMAFVPTAIPSPHQ	558
Qy	795	ORLOYTRKLLMEQQAFLVSSAPRYLYLASNSNKGHGRGRIOMLSAPAGEPLONS	855
Db	559	ORLOYTRKQLTEQQAAPLGGASPRYLYLASKOSNKGHGRGRIQTVSPASGAPMPONS	618
Qy	855	SMARCFSWERYQLAVTORKEEPESSSVFNONDPAVTPFSDIINNETHGKDLVAMV	914
Db	619	PMERAFMSGRGRLAITQKEKEPESSSVFNONDPAVTPFSDIINNETHGKDLVAMV	678
Qy	915	AGFLTHIPHAEDIPNTVTVGNVGFFLRPNFFDEDPSTYSADSIYFRGDDAGACEVNP	974
Db	679	AGFLTHIPHAEDIPNTVTVGNVGFFLRPNFFDEDPSPADSISYFRGDDAGACEINPL	738
Qy	975	ACLPQAACAPDLPFAFSGGF	995
Db	739	ACLPQAATCAPDLPVFSHG	759
RESULT 5			
AOC2_HUMAN			
ID	AOC2_HUMAN	STANDARD:	PRT: 756 AA.
AC	075105; 075105; 000120; 09UNO;		
DT	15-JUL-1999 (Rel. 38, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Retina-specific copper amine oxidase precursor (EC 1.4.3.6) (RAO)		
GN	(Amine oxidase [copper-containing]).		
GC	AOC2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Retina;		
RA	MEDLINE=97237047; PubMed=9119395;		
RA	Imamura Y., Kubota R., Wang Y., Asakawa S., Kudoh J., Mashima Y.,		
RA	Oguchi Y., Shimizu N.;		
RT	"Human retina-specific amine oxidase (RAO): cDNA cloning, tissue		
RT	expression, and chromosomal mapping.";		
RL	Genomics 40:277-283(1997).		
RL	[2]		
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.		
RA	MEDLINE=98390194; PubMed=9722954;		
RA	Imamura Y., Noda S., Mashima Y., Kudoh J., Oguchi Y., Shimizu N.;		
RT	"Human retina-specific amine oxidase: genomic structure of the gene		
RT	(AOC2), alternatively spliced variant, and mRNA expression in		
RL	retina.";		
RL	Genomics 51:293-298(1998).		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Retina;		
RA	"Zhang X., McIntire W.S.;		
RT	"Human copper-containing amine oxidases.";		
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: MAY BE A CRITICAL MODULATOR OF SIGNAL TRANSMISSION IN		
CC	RETINA. POSSIBLY BY DEGRADING THE BIOGENIC AMINES DOPAMINE,		
CC	HISTAMINE, AND PUTRESCINE.		
CC	-1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +		

```

CC      H(2)O(12).
CC      -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC      -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: A LONG FORM (SHOWN
CC      HERE) AND A SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: RETINAL-SPECIFIC.
CC      -1- PM: Topaquinine (TPO) is generated by copper-dependent
CC      autooxidation of a specific tyrosyl residue (by similarity).
CC      -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see license@isb-sib.ch).
CC      -----
DR      EMBL; AB012943; BAA32590.1; -.
DR      EMBL; AB012942; BAA32590.1; JOINED.
DR      EMBL; AB012943; BAA32589.1; -.
DR      EMBL; AB012942; BAA32589.1; JOINED.
DR      EMBL; D88213; BAA19001.1; -.
DR      EMBL; AF081363; AAC9345.1; -.
DR      Genew; HGNC:549; AOC2.
DR      MIM; 602268; -.
DR      InterPro; IPR000269; CuNH_oxidase.
DR      Pfam; PF01179; Cu_amine_oxid; 1.
DR      Pfam; PF02727; Cu_amine_oxidn2; 1.
DR      Pfam; PF02728; Cu_amine_oxidn3; 1.
DR      PRINTS; PR00766; CUDAOXIDASE.
DR      PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.
DR      PROSITE; PS01165; COPPER_AMINE_OXID_2; 1.
KW      Oxidoreductase; Copper; TPO; Alternative splicing; Glycoprotein;
KW      Signal; Metal-binding.
FT      SIGNAL          1..32
FT      CHAIN           33..756
FT      MOD_RES         465..465
FT      METAL           516..516
FT      METAL           518..518
FT      METAL           680..680
FT      BINDING         669..669
FT      CARBOHYD        133..133
FT      CARBOHYD        198..198
FT      CARBOHYD        226..226
FT      CARBOHYD        588..588
FT      CARBOHYD        662..662
FT      VARSPLIC        599..625
FT      CONFLICT        181..181
FT      CONFLICT        215..218
FT      CONFLICT        221..222
FT      CONFLICT        610..610
SQ      SEQUENCE       756 AA; 83793 MW; 4E3B7317E6DACA66 CRC64;
Query Match          47.7%; Score 2556; DB 1; Length 756;
Best Local Similarity 65.8%; Pred. No. 1.8e-183;
Matches 480; Conservative 88; Mismatches 158; Indels 4; Gaps 2
Db      207

```

```

OY 508 SLAOLEAFOEAGLVNVLIPNGTSGWSLSPVPGPAPPLQYPOGPRESVGSGRYAS 567
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 267 DLGLEBREFKGRLEVRVPLPPNGASSLRSRNSPGBPLPLOSPOSGQTSVGNLVS 326
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 568 SLMTFSGLAGFSGPRIFDVRFQGERLYVEISLOALAIYGGNSPAAATTTRYVGGFGMG 627
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 327 SLMSFTGHHGFGSLRIFDVRFQGERIAYEVSGVCEVSGADSPKMTLTRYLDSSTGLG 386
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 628 KYTTPPLRGVDCPLATYVDMHFLLESQAPRTIDAFVCEQONGLPLRRHSDLYSHYF 687
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 387 RNSRGLVWGVDCPYQATMVVHILVKGAVOLLPGAVCEVEAGGLPLRRHNYLQNHFY 446
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 688 GGLAETLVVRSMTLTNDVDMTFVPHPSGAIRFATYISAFIFGATG--KYGNO 745
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 447 GGLASSALVVRSSVSGNYDIMPVLYPNALGCRVHATGYINTAFKGGEGGLFGRN 506
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 746 VSEHTLCTVHTSHAFVDDVAGLENNVVAEDVFPVMAVPSPEHQLQVTRYKLE 805
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 507 VGEHVLTGVTHAHFGLDLDDVAGLKNMVAEDVFKFPAAPMPEHMLQAPQLTRYGLG 566
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 806 MEEGAALVGSATPRYLVLASNNKMGHPRGVRIOMLSFGEBELPONSSAARGFSNERY 865
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 567 KEDLTARSLGSPRLPRYLVLASNOTNMGHORGVRIQIHSPIGIDIPLESDEMERALSNGRY 626
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 866 QLAAYTORKEEPPSSSVFNQNDPMPAPTVDSDFINNETIAGKDLVAVMTAGFLHIPHED 925
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 627 QLVYTORKEEPPSSSVSYIHNQNDIPTVTFADFINNETILGEDLVAVMTAGFLHIPHED 686
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 926 IPNTVTVNGNGVGFLLRPYNEFDEDPSTYSADSIYFRCGDQAGACEVNPACLPOAACAP 985
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 687 IPNTVTVNGNGVGFLLRPYNEFDEDPSTYSADSIYFRCGDQAGACEVNPACLPOAACAP 985
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 986 DLPAFSGHF 995
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 747 DLPPFSYHGF 756
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |

RESULT 6
ABP_RAT ID STANDARD: PRT: 746 AA.
AC P3633;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Amloride-sensitive amine oxidase [copper-containing] precursor
  (EC 1.4.3.6) (diamine oxidase) (DAO) (amloride-binding protein)
DE (ABP) (Histaminase).
GN ABP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerogomathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Colon, and Lung;
RA MEDLINE=93387321; Pubmed=8375402;
RX Linguigla E., Renard S., Voilley N., Waldmann R., Chassande O.,
  Lazdunski M., Barbry P.;
  "Molecular cloning and functional expression of different molecular
  forms of rat amloride-binding proteins.";
  Eur. J. Biochem. 216:679-687(1993).
RL - FUNCTION: CATALYZES THE DEGRADATION OF COMPOUNDS SUCH AS
  PUTRESCINE OR HISTAMINE.
CC - CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
  H(2)O(2).
CC - COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC - SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC - SUBCELLULAR LOCATION: Extracellular.
CC - PTM: Topaquinine (TPO) is generated by copper-dependent
  autoxidation of a specific tyrosyl residue (by similarity).
CC - SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X73911; CA52116.1; -.
DR EMBL: X73912; CA52117.1; ALT_INIT.
DR PIR: S34656; S34656.
DR PIR: S36847; S36847.
DR InterPro: IP000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid.
DR Pfam: PF02727; Cu_amine_oxid2.
DR Pfam: PF02728; Cu_amine_oxid3.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID.
DR PROSITE: PS01165; COPPER_AMINE_OXID.
KW Signal; Glycoprotein; Oxidoreductase; Copper; Heparin-binding; TPO;
KW Metal-binding.
FT CHAIN 1..22
FT MOD_RES 456 456
FT METAL 505 505
FT METAL 507 507
FT METAL 670 670
FT BINDING 563 570
FT BINDING 659 659
FT CARBOHYD 110 110
FT CARBOHYD 269 269
FT CARBOHYD 533 533
FT CARBOHYD 740 740
FT SEQUENCE 746 AA; 85021 MW; 6C564D04D07BCB2 CRC64;

Query Match 26.1%; Score 1398.5; DB 1; Length 746;
Best Local Similarity 40.3%; Pred. No. 8.e-97;
Matches 294; Conservative 127; Mismatches 266; Indels 43; Gaps 13;

OY 291 QSLFADLSREELTAANKRFLTORLGPGLVDAQAQPSDNCVFSYEQLPKKAALAHLD 350
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 29 KAVFADLSPEIKAVHSFLMNRBELGLQPSKEPTLAKNSVFLIMLLPKKKHYLKLDE 88
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 351 GSPPAEALAIYFSGNQPOPNVSELYVGPLPHSYRDVTEHNGRPLRYHRPVLP 410
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 89 GRKPNREARVIFFGADVPNVEFAVGPLPRYVIRALS-PPGHNLSSSPISITAE 147
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 411 YLDIDQIFNRELPOASGLHH-----CCFYKRGRLVTMTTAPRGLOSGDRTWFG 463
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 148 Y----DLIYH-TLKRATMPPLHGFLLDTGTFGLGCDRCRLFTVYARGVASGGRBMFI 202
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 464 LYVINSAGGFLHHVGLLELVNKRALDPAKWTLOKVFYQGRYDLSLAOLEAFOEAGLVN 523
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 203 VQRYVE--GYFLHPTGLTEILLDHGSTDVQDMRVQLVNGKFNPNPELARKYAVGEVD 260
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 524 VLIRD--NGTSGWSLSKSPV-----PPGAPRLQFPQGPRESSVQSGRASS 568
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 261 VLEDDPLPNKTE---KPLFSSYKRGGEFHVNVNAGPVVQSPGRKLEENLYLG 315
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 569 LMTFSGLAGFSGPRIFDVRFQGERLYVEISLOALAIYGGNSPAAATTTRYVGGFGMG 628
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 316 GMSYSTLRSSSGQILNVLFGEGERVAYEVQGAVALVGHGHPAGQMTYIDVQWGLG 375
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 629 YTTPLTGVDCPYLATYVDMHFLLESQAPRTIDAFVCEQONGLPLRRHSDLYS--H 685
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 376 VTHELAGIDCPERATFLDAFHYDSDGPHYPRALCLFEPTGVPPLRRHFNNSFKGFGN 435
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 686 YFGLAFTLVVRSMTLTNDVDMTFVPHPSGAIRFATYISAFIFGATG--KYGNO 745
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
DB 436 FYAGLKVTLVLRKTSVYVINDIYWDIFYSNGVEMKHAATGVHATFTTPBSGLRGT 495
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |
OY 746 VSEHTLCTVHTSHAFVDDVAGLENNVVAEDVFPVMAVPSPEHQLQVTRYKLE 805
    | | | | | : | | : | | : | | | | | : | | | | | : | | | | | : |

```

Db 496 LQTHLLGNIHVLHYRVDMDVAGTKNSFQITMLKLENTLNWSPSHSLVQFTLQTOYS 555
 QY 806 MEEQAFVLSATPRYLTLASNNKMGHPRGYRIQMLSFAGEPLPONSSMARGSWERY 865
 Db 556 QEHQAFVFGQGLPKYLLFSSPOKCKMGRRSRYLQISMARQVLPFGQGEARATWARY 615
 QY 866 QLVATQKEEPPSSSVFNQNDPAPTYDFSDFT-NNEFIACKDLVAVYTAGFLHHPAE 924
 Db 616 PLATKYRESERYSSSLYNQNDPMDPVYFEFEFLRNNEDEDELVAVYTAGFLHHPSE 675
 QY 925 DIPNVTGNGVGFRLRPNEFDEPSPFSADSIYFRGQDAGACEVNPACLPQAAACA 984
 Db 676 DVPNATGNSYGFULRPFNPEDPSPSLASRDTVIY-WFQKGLNRVQ--RMIPEDRCL 732
 QY 985 PDLPAFSGHG 994
 Db 733 VS-PPFSYNG 741

RESULT 7
 ABP_HUMAN STANDARD: PRT; 751 AA.
 ID AC P19801: 016683; 016684;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Amiloride-sensitive amine oxidase [copper-containing] precursor
 DE (EC 1.4.3.6) (diamine oxidase) (DAO) (Amiloride-binding protein)
 DE (ABP) (Histaminase).
 GN ABPI OR DAO1 OR AOCL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94237856; PubMed-8182053;
 RA Chassande O., Renard S., Barbry P., Lazdunski M.;
 RT "The human gene for diamine oxidase, an amiloride binding protein.
 RT Molecular cloning, sequencing, and characterization of the promoter.";
 RL J. Biol. Chem. 269:14484-14489(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE-Placenta;
 RX MEDLINE-96113540; PubMed-8595053;
 RA Zhang X., Kim J., McIntire W.S.;
 RT "cDNA sequences of variant forms of human placenta diamine oxidase.";
 RL Biochem. Genet. 33:261-268(1995).
 RN [3]
 RP PRELIMINARY SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX TISSUE-Kidney;
 RX MEDLINE-91017502; PubMed-2217167;
 RA Barbry P., Champe M., Chassande O., Munemitsu S., Champigny G.,
 RA Linqueglia E., Maes P., Frelin C., Tartar A., Ullrich A.,
 RA Lazdunski M.;
 RT "Human kidney amiloride-binding protein: cDNA structure and functional
 RT expression.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7347-7351(1990).
 RN [4]
 RP SEQUENCE OF 20-39, AND CHARACTERIZATION.
 RX TISSUE-Placenta;
 RX MEDLINE-94193685; PubMed-8144586;
 RA Novotny W.F., Chassande O., Baker M., Lazdunski M., Barbry P.;
 RT "Diamine oxidase is the amiloride-binding protein and is inhibited by
 RT amiloride analogues.";
 RL J. Biol. Chem. 269:9921-9925(1994).
 CC -1- FUNCTION: CATALYZES THE DEGRADATION OF COMPOUNDS SUCH AS
 CC PUTRESCINE, HISTAMINE, SPERMINE, AND SPERMIDINE. SUBSTRANCES
 CC INVOLVED IN ALLERGIC AND IMMUNE RESPONSES, CELL PROLIFERATION,
 CC TISSUE DIFFERENTIATION, TUMOR FORMATION, AND POSSIBLY APOPTOSIS.
 CC PLACENTAL DAO IS THOUGHT TO PLAY A ROLE IN THE REGULATION OF THE
 CC FEMALE REPRODUCTIVE FUNCTION.
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: PLACENTA AND KIDNEY.
 CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X78212; CAA55046.1; -;
 CC EMBL: U11862; AAC50270.1; -;
 CC EMBL: U11863; AAB60381.1; -;
 CC EMBL: M55602; AAA58358.1; ALT_SEQ.
 CC PIR: A38276; A38276.
 CC PIR: S42495; S42495.
 CC Genew: HGNC:80; ABPI.
 CC MIM: 104610; -;
 CC InterPro: IPR00269; CUNH_Oxidase.
 CC Pfam: PF01179; Cu_amine_oxid; 1.
 CC Pfam: PF02727; Cu_amine_oxidn2; 1.
 CC Pfam: PF02728; Cu_amine_oxidn3; 1.
 CC PRINTS: PR00766; CUOXIDASE.
 CC PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 CC PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 CC Signal: Glycoprotein; Oxidoreductase; Copper; Heparin-binding; TPQ;
 CC Alternative splicing; Polymorphism; Metal-binding.
 CC KW SIGNAL 1 19
 CC CHAIN 20 751
 CC FT METAL 391 391
 CC FT MOD_RES 461 461
 CC FT METAL 510 510
 CC FT METAL 512 512
 CC FT METAL 675 675
 CC FT BINDING 568 575
 CC FT BINDING 664 664
 CC FT CARBOHYD 110 110
 CC FT CARBOHYD 168 168
 CC FT CARBOHYD 538 538
 CC FT CARBOHYD 745 745
 CC FT VARSPLIC 619 619
 CC FT VARIANT 645 645
 CC FT CONFLICT 28 28
 CC FT CONFLICT 332 332
 CC FT SEQUENCE 751 AA; 85363 MW; 1B8B6595C44232DA CRC64;
 CC
 CC Query Match 26.0%; Score 1394; DB 1; Length 751;
 CC Best Local Similarity 39.8%; Pred. No. 1, 9e-96;
 CC Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;
 CC
 CC 285 PMTHPGOSOLFADLSRELAVNRFLRQGPGLVDAQAARSPNCFVSVELQPPKAA 344
 CC 23 PGLTPKAGVSDISNOELAAVHSLFMSKELRQPSSTTMKNTYFLEMLPKRYH 82
 CC 345 LAHLDRGSPPEALAIVEFGROPVSELVVPLPHPSYMDVTVRHHGGLPYHRR 404
 CC 83 LRFIDKGERPVRVARAVIFFGDEHPNTEFAVGPLPCGYMALR-PPRGVSSWASR 141
 CC 405 PVLFQELLDQMFNLELQASGLNH-----CCFYKHGRNLTMTTAPRGLOSGD 457
 CC 142 PISTAEV-----ALVYHTLOEATKPLHGFLLNTTGFSDCHDCLAFVDAAPGASGQ 196
 CC 458 RATWFGLYYNISAGFLHVGLELLVNHKALDPAWTIQKVYQGRYDSLAOLEAOF 517

```

Db 197 RRSLLIQRVE--GYFLHPTGLLELVHGSTDAGHNAVEQVYNGKRYGSPEELARKYA 254
QY 518 AGLVNVLIPDNGTGGMSLSKSPVPR-----GPAP-----PLQFYQGRFSYQG 562
Db 255 DGEVDVVLLEDPPLFGGCHDSTPEPLPSSHKPRGDPSPPLHVSQPLVQHGFRFLREG 314
QY 563 SVVASSLMTSFGIGAFSGRIEDVAFQGERLYVEISLOELALTYGNSPAAMTTRYVDG 622
Db 315 NAVLYGMSAFRLRSSSGLOVLNVHGERIANVEVQEAVALYGHGTAGMGTITLDV 374
QY 623 GFQMGKTTPLTRGVDCPYLATYVDHMLLESQAPKTIRDAFCVEQNOGLPLRRHNSDL 682
Db 375 GWGLGSYTHIELAGIDCEPETAFELDTFHYDDADDPVHYPRALCLFEMPTGVPLRHNSN 434
QY 683 YS--HFGGLAETVLVYRSMSTLLANDYWDYFHSGLAIEIFYATGYSSAFLEGAT 739
Db 435 FKGGFNYYAGLKGVALLRTSTVYNDYITWDFLEPNGVEMAKHATGYVHAFTYPEG 494
QY 740 GKYGNQVSEHTLGTHTSHAFKVDLDVAGLENNVMAEDVFMVMAVMSPEHOLQY 799
Db 495 LRHGTRLHTHLIGNHHLVHYRDLVAGTKNSFQLOKLENTITPWSFRHVVQPTL 554
QY 800 TRLLLEMEQAAFLVGSATPRYLYLASNSHNGKHPRGYRIQLSFAGEPLPQNSMARG 859
Db 555 EQQYVSERQAAFRFKRLKPLKYLFTSPQENPWGKRSYRLQIHSMADQVLPQWQEQ 614
QY 860 FSWERYOLATQKREKPESSSVFNQNDPAFVYDSDFI--NNETIGKDLVAANTAGFL 918
Db 615 ITARIPLATATKRESLCSSTIYHNDPMWPRVYEQFQHNENINEDLVAAVTGVL 674
QY 919 HIPHADIPNTVYVNGVGFELRPYNFEDDPSFYSADSIYFRDODAGACEVPLACL 978
Db 675 HIPSIEDIPATATPGNSVGLRLPFPNFPEDPSLASMDIYV--WPRNGENYV--RWIP 731
QY 979 QAAACAPDLPAFESHG 994
Db 732 EDRDCSMP--PPFSYNG 746

RESULT 8
AOC3_RAT
ID AOC3_RAT STANDARD: PRT; 321 AA.
AC 008590;
DC 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 39, Last sequence update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1) (V997) (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Adipocyte;
RX MEDLINE=97238878; PubMed=9083076;
RA Morris N.J., Ducret A., Aebersold R., Ross S.A., Keller S.R.,
RL Lienhard G.E.;
RT "Membrane amine oxidase cloning and identification as a major protein in the adipocyte plasma membrane."
RT J. Biol. Chem. 272:9388-9392(1997).
RL (2)
RP SEQUENCE OF 1-19.
RC TISSUE=Adipocyte;
RX MEDLINE=96010417; PubMed=8520629;
RA Jochen A., Guven S., Hays J.;
RT "The major integral membrane glycoprotein in adipocytes is a novel 200-kDa heterodimer."
RT Mol. Membr. Biol. 12:277-281(1995).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN I-SELECTIN-

```

```

CC INDEPENDENT FASHION. HAS A MONOMERIC OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) + H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- Ptm: Topaquinine (TPQ) is generated by copper-dependent autooxidation of a specific tyrosyl residue (by similarity).
CC -1- Ptm: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U72632; AAC5189.1; -.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUOAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; PARTIAL.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; PARTIAL.
KW Oxidoreductase; Copper; TPQ; Glycoprotein; Transmembrane;
FT Signal-anchor; Cell adhesion; Metal-binding.
FT INIT_MET 0 0
FT DOMAIN 1 5
FT TRANSMEM 6 26
FT FT CYTOPLASMIC (POTENTIAL).
FT FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 27 >321
FT CARBOHYD 136 136
FT CARBOHYD 231 231
FT CARBOHYD 293 293
FT NON_TER 321 321
SQ SEQUENCE 321 AA; 35116 MW; 6B1F294253A1D6F CRC64;

Query Match 24.5%; Score 1315; DB 1; Length 321;
Best Local Similarity 82.7%; Pred. No. 4.5e-91;
Matches 244; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

QY 263 QCGDGGSPQLPCHPSVSPAQWTHPGSQSLPADISREELAVMRFQRLGRLGVDA 322
Db 27 RSDSDGRLSQPLCPVLPSPVQPTGSGQSPFADISPEELTAVMSFLNHLRPLGVDA 86
QY 323 QARPSPNCVSVLEQLPRAKAAALAHLDGSPPARALATVFGROPVSELYGCLP 382
Db 87 QARPSPNCVSVLEQLPRAKAAALAHLDGSPPARALATVFGROPVSELYGCLP 146
QY 383 HPSYMDVYVERHGRLPYHRRPVLFOEYLDIDOMTFNRELPOASGLHHCFFYKGRN 442
Db 147 HPSYMDVYVERHGRLPYHRRPVLFOEYLDIDOMTFNRELPOASGLHHCFFYKGRN 206
QY 443 LVYMTAPRLQSGDPAATWGLYNSGAGFLAHVGLLVNKKALDPAKMTIQVQY 502
Db 207 LKMTTAPRLQSGDPAATWGLYNSGAGFLAHVGLLVNKKALDPAKMTIQVQY 266
QY 503 GRYYSIAOLEAOFEGALVNVVLIPDNGTGGMSLSKSPVPPGAPRLQYPOGR 557
Db 267 GRYYSIAOLEAOFEGALVNVVLIPDNGTGGMSLSKSPVPPGAPRLQYPOGR 321

RESULT 9
GT26_SCHJA
ID GT26_SCHJA STANDARD: PRT; 218 AA.
AC P08515;
DC 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DR 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (Sj26 antigen) (GST class-alpha).

```


Dh 62 DVKLTQSMALIRYIADKHNMLGGCPKERAELISLEGALIDIRGVSRVAYNKEFTLKYD 121
Oy 158 FLSKLPEMLKMFEDRLSKRTYLNGLDVTHTPPDMLYDALDVLYXMDPCLDAFPKLVSPK 217
Db 122 FLNQLPGLKMFEDRLSKRTYLNGLDVTHTPPDMLYDALDVLYXMDPCLDAFPKLVSPK 181
Oy 218 RLEALPQIDKYLKSSKYIAMPLOGQWATFGGDDHPK 254
Db 182 RIENLPPIKYNLNSRYIKWPLQGSATFGGDDAPK 218

RESULT 11

GT26_SCHMA STANDARD; PRF: 218 AA.
AC P15964;

Dt 01-APR-1990 (Rel. 14, Created)
Dt 01-APR-1990 (Rel. 14, Last sequence update)
Dt 01-NOV-1995 (Rel. 32, Last annotation update)
De Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SW26/1 antigen) (GST class-alpha).
Os Schistosoma mansoni (Blood fluke).
Oc Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae;
Ox Schistosomatidae; Schistosomatidae; Schistosoma.
NCBI_TaxID=6183;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Puerto Rican.
RX MEDLINE=90348716; PubMed=2385266.
RA Trottein F., Kieny M.P., Verweide C., Torpier G., Pierce R.J.,
RT "Molecular cloning and tissue distribution of a 26-kilodalton
RT Schistosoma mansoni glutathione S-transferase.";
RL Mol. Biochem. Parasitol. 41:35-44(1990).
RN [2]
RP SEQUENCE OF 8-218 FROM N.A.
RC STRAIN-Puerto Rican.
RX MEDLINE=90271935; PubMed=1693415;
RA Hedke K.J., Davern K.M., Wright M.D., Ramos A.J., Mitchell G.F.;
RT "Comparison of the cloned genes of the 26- and 28-kilodalton
RT glutathione S-transferases of Schistosoma japonicum and Schistosoma
RT mansoni.";
RL Mol. Biochem. Parasitol. 40:23-34(1990).

CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- TISSUE SPECIFICITY: TEGUMENT AND IN SUBTEGUMENTARY PARENCHYMAL
CC CELLS. GST 26 MAY BE ACTIVELY EXCRETED BY ADULT WORMS.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST THREE ISOENZYMES OF GST IN
CC S. MANSONI.

CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL: M31106; AAA29888.1; -
CC DR EMBL: M26913; AAA29889.1; -
CC DR PIR: A45523; A45523.
CC DR HSSP: P08515; 1GTA.
CC DR InterPro: IPR004046; GST_Cterm.
CC InterPro: IPR004045; GST_Nterm.
CC Pfam: PF00043; GST_C; 1.
CC Pfam: PF02798; GST_N; 1.
CC DR Transferase; Antigen; Multigene family.

SQ SEQUENCE 218 AA; 25401 MW; 061A6548A842D6E8 CRC64;
Query Match 17.2%; Score 922; DB 1; Length 218;
Best Local Similarity 80.2%; Pred. No. 6.9e-62;
Matches 174; Conservative 15; Mismatches 28; Indels 0; Gaps 0;

Oy 38 SPILGWIKIGLVQPTRLLEYLEEKEHYERDEGQWRNKKFELGEPNLPYYIDG 97
Db 2 AFKFGYWKYKGLVQPTRLLEHLEETYEERADRNEDAMSNDRKFKLGLEFPNLPYYIDG 61
Oy 98 DVKLTQSMALIRYIADKHNMLGGCPKERAELISLEGALIDIRGVSRVAYNKEFTLKYD 157
Db 62 DVKLTQSMALIRYIADKHNMLGGCPKERAELISLEGALIDIRGVSRVAYNKEFTLKYD 121
Oy 158 FLSKLPEMLKMFEDRLSKRTYLNGLDVTHTPPDMLYDALDVLYXMDPCLDAFPKLVSPK 217
Db 122 FLNQLPGLKMFEDRLSKRTYLNGLDVTHTPPDMLYDALDVLYXMDPCLDAFPKLVSPK 181
Oy 218 RLEALPQIDKYLKSSKYIAMPLOGQWATFGGDDHPK 254
Db 182 CIEDLPQIKYNLNSRYIKWPLQGSATFGGDDAPK 218

RESULT 12

GT29_FASHE STANDARD; PRF: 220 AA.
AC P56598;

Dt 15-DEC-1998 (Rel. 37, Created)
Dt 15-DEC-1998 (Rel. 37, Last sequence update)
Dt 15-DEC-1998 (Rel. 37, Last annotation update)
De Glutathione S-transferase 26 kDa 1 (EC 2.5.1.18) (GST1) (FHL) (GST class-alpha).
Os Fasciola hepatica (Liver fluke).
Oc Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Ox Echinostomida; Echinostomata; Fasciolidae; Fasciola.
NCBI_TaxID=6192;

RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92153506; PubMed=1740183;
RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
RT transferases of Fasciola hepatica.";
RL Exp. Parasitol. 74:232-237(1992).
RN [2]
RP EXP. PARASITOL. 74:232-237(1992).

CC ERRATUM.
CC MEDLINE=94039664; PubMed=8224094;
CC Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
CC Exp. Parasitol. 77:385-385(1993).
CC RN [3]
CC RP SEQUENCE OF 22-220 FROM N.A.

CC RA Cramer S.L.
CC Patient number W09008819, 09-AUG-1990.
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- SUBUNIT: HOMODIMER.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL: A00993; CAA00118.1; -
CC DR HSSP: P31670; 1FHE.
CC DR InterPro: IPR004046; GST_Cterm.

DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase; Antigen; Multigene family.
 FT INIT MET 0 BY SIMILARITY.
 FT CONFLICT 22 22 DP -> V (IN REF. 3).
 FT CONFLICT 110 111 A -> P (IN REF. 3).
 FT CONFLICT 189 189
 SQ SEQUENCE 220 AA: 25598 MW: 2789150B/5D10IF CRC64;
 Query Match 13.4%; Score 719.5; DB 1; Length 220;
 Best Local Similarity 59.2%; Pred. No. 1e-46;
 Matches 199; Conservative 34; Mismatches 54; Indels 1; Gaps 1;
 QY 41 LGYWKIKGLVQPTRLLEYLEEKYEHLXERDEGDKMKNKFEGLGEPNLPYIDGVK 100
 DB 4 LGYWKIKGLVQPTRLLEYLEEKYEHLXERDEGDKMKNKFEGLGEPNLPYIDGVK 62
 QY 101 LTQSMATIRYIADKHNMLGSGPKRAEISMLEGAVLDIRGVSRVSKOPELTKVDFLS 160
 DB 63 LTQSMATIRYIADKHNMLGSGPKRAEISMLEGAVLDIRGVSRVSKOPELTKVDFLS 122
 QY 161 KLPEMLKMFEDRLSKHTYLNQDHTHPDMLYDALDVLYLMDPCLDAFPKLVSEFKKRIE 220
 DB 123 DLPTTKMMSDFLGKKNPILKRTSVSHVDPMYELDAIRYLEPHCLDHPMLQPMRSRIE 182
 QY 221 AIPQIDKYLKSSKIYAMPLOGMQATFGGDDHPKRSQSQ 258
 DB 183 ALPKIKAYMESNRFIKWPPLNGMHAQFGGDAPPSHEKK 220
 RESULT 13
 GT28_FASHE STANDARD; PRT: 217 AA.
 ID AC P31670;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GST47) (FH47)
 DE (GST class-alpha).
 OS Fasciola hepatica (liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fasciolidae; Fasciola.
 OX NCBI_TaxID=6192;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92155306; PubMed=1740183;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 transferases of Fasciola hepatica."
 RL Exp. Parasitol. 74:232-237(1992).
 RN (2)
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN (3)
 RP SEQUENCE OF 7-105 FROM N.A.
 RA Cramer S.;
 RL Patent number WO9008819, 09-AUG-1990.
 RN (4)
 RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
 RX MEDLINE=98035725; PubMed=9367777;
 RA Rosjohn J., Fell S.C., Wilce M.C.J., Sexton J.L., Spithill T.W.,
 RA Parker M.W.;
 RT "Crystallization, structural determination and analysis of a novel
 parasite vaccine candidate: Fasciola hepatica glutathione
 S-transferase."
 RL J. Mol. Biol. 273:857-872(1997).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO

CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M77681; AAA29140.1; -;
 DR EMBL: A00996; CA00121.1; -;
 DR PDB: 1FHE; 29-JUL-98.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase; Antigen; Multigene family; 3D-structure.
 FT INIT MET 0
 FT CONFLICT 65 65 T -> I (IN REF. 3) f
 FT CONFLICT 102 105 RIGF -> FEEL (IN REF. 3).
 SQ SEQUENCE 217 AA: 25281 MW: 0FB8B8FE63029E03 CRC64;
 Query Match 13.2%; Score 708; DB 1; Length 217;
 Best Local Similarity 58.2%; Pred. No. 7.3e-46;
 Matches 124; Conservative 37; Mismatches 52; Indels 0; Gaps 0;
 QY 41 LGYWKIKGLVQPTRLLEYLEEKYEHLXERDEGDKMKNKFEGLGEPNLPYIDGVK 100
 DB 4 LGYWKIKGLVQPTRLLEYLEEKYEHLXERDEGDKMKNKFEGLGEPNLPYIDGVK 63
 QY 101 LTQSMATIRYIADKHNMLGSGPKRAEISMLEGAVLDIRGVSRVSKOPELTKVDFLS 160
 DB 64 LTQSMATIRYIADKHNMLGSGPKRAEISMLEGAVLDIRGVSRVSKOPELTKVDFLS 123
 QY 161 KLPEMLKMFEDRLSKHTYLNQDHTHPDMLYDALDVLYLMDPCLDAFPKLVSEFKKRIE 220
 DB 124 ELPTKLMMSDFLGKKNPILKRTSVSHVDPMYELDAIRYLEPHCLDHPMLQPMRSRIE 183
 QY 221 AIPQIDKYLKSSKIYAMPLOGMQATFGGDDHPKRSQSQ 253
 DB 184 ALPKIKAYMESNRFIKWPPLNGMHAQFGGDAPPSHEKK 216
 RESULT 14
 GT28_FASHE STANDARD; PRT: 217 AA.
 ID AC P31671;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GST7) (FH7) (GST
 DE class-alpha).
 OS Fasciola hepatica (liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fasciolidae; Fasciola.
 OX NCBI_TaxID=6192;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92155306; PubMed=1740183;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 transferases of Fasciola hepatica."
 RL Exp. Parasitol. 74:232-237(1992).
 RN (2)
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;


```

RN Exp. Parasitol. 77:385-385(1993).
RA [3]
RP SEQUENCE OF 8-217 FROM N.A.
RA Cramerl S.;
RL Patent number WO9008819, 09-AUG-1990.
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: M77680; AAA29141.1; -
DR EMBL: A00994; CAA00119.1; -
DR HSSP: P31670; 1FHE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C_1.
DR Pfam: PF02798; GST_N_1.
DR Transferase; Antigen; Multigene family.
DR INIT_MET 0 BY SIMILARITY.
FT CONFLICT 35 NDREKMG -> MGRNGMA (IN REF. 3).
FT CONFLICT 188 IKEYKSR -> SKYMSA (IN REF. 3).
SQ SEQUENCE 217 AA; 25196 MW; 0099E1F59E49A49E CRC64;

Query Match 12.9%; Score 690; DB 1; Length 217;
Best Local Similarity 58.7%; Pred. No. 1.9e-44;
Matches 125; Conservative 30; Mismatches 58; Indels 0; Gaps 0;

OY 41 LGYWKIKGLVPTRLLELEYEEHLYERDEGDKRNRKKELEGLFEPNLPYTIIDGVK 100
DB 4 LGYWKIRGLQGVRLLELEYEEHLYERDEGDKRNRKKELEGLFEPNLPYTIIDCK 63
OY 101 LTOSMAIIRYIADKHNMLGSPKRAEISMLGAVLDIRGYRSIYKDFELKVDLFS 160
DB 64 LTQSAVAINRYIADKHNMLGSPKRAEISMLGAVLDIRGYRSIYKDFELKVDLFS 123
OY 161 KLPEMLKMFEDRLSHKTYLNGDHTVHPDMLYDALDVLYMDPCLDAFPKLVSEKRIE 220
DB 124 GLPTLKMNSDFLGRHLYLTSVSHVDFMYEALDCIRYLAPOCLDEFPKLEKRSRIE 183
OY 221 AIPQIDKYLKSKYIAMPLOGMOAFFGGGDHP 253
DB 184 DLPKIKAYMESEKRIKMPPLNSHVSFSGGDHP 216

RESULT 15
GT26_FASHE STANDARD: PRT: 217 AA.
AC P30112:
DT 01-APR-1993 (Rel. 25, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 51 (EC 2.5.1.18) (GSTS1) (FHS1)
DE (GST class-alpha).
OS Fasciola hepatica (Liver fluke).
OC Echinostomida; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomidae; Echinostomata; Fasciolidae; Fasciolidae; Fasciola.
RN NCBI_TaxID-6192;
RP SEQUENCE FROM N.A.
RX MEDLINE-92155306; Pubmed-1740183;

```

```

RA Panaccio M., Wilson L.R., Cramerl S.L., Wajffels G.L., Splithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
RT transferases of Fasciola hepatica.";
RL Exp. Parasitol. 74:232-237(1992).
RN [2]
RP ERRATUM.
RX MEDLINE-94039664; Pubmed-8224094;
RA Panaccio M., Wilson L.R., Cramerl S.L., Wajffels G.L., Splithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 9-217 FROM N.A.
RX MEDLINE-93228188; Pubmed-7692383;
RA Moro A., Rodriguez-Molina J.R., Hillyer G.V.;
RT "Sequence analysis of a Fasciola hepatica glutathione S-transferase
RT cDNA clone.";
RL Am. J. Trop. Med. Hyg. 48:457-463(1993).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: M77682; AAA29141.1; -
DR HSSP: P31670; 1FHE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C_1.
DR Pfam: PF02798; GST_N_1.
DR Transferase; Antigen; Multigene family.
DR INIT_MET 0 BY SIMILARITY.
FT CONFLICT 83 T -> S (IN REF. 3).
FT CONFLICT 134 N -> D (IN REF. 3).
FT CONFLICT 146 P -> T (IN REF. 3).
SQ SEQUENCE 217 AA; 25242 MW; EFC88FE7454ECC26 CRC64;

Query Match 12.9%; Score 689; DB 1; Length 217;
Best Local Similarity 58.7%; Pred. No. 1.9e-44;
Matches 125; Conservative 33; Mismatches 55; Indels 0; Gaps 0;

OY 41 LGYWKIKGLVPTRLLELEYEEHLYERDEGDKRNRKKELEGLFEPNLPYTIIDGVK 100
DB 4 LGYWKIRGLQGVRLLELEYEEHLYERDEGDKRNRKKELEGLFEPNLPYTIIDCK 63
OY 101 LTOSMAIIRYIADKHNMLGSPKRAEISMLGAVLDIRGYRSIYKDFELKVDLFS 160
DB 64 LTQSAVAINRYIADKHNMLGSPKRAEISMLGAVLDIRGYRSIYKDFELKVDLFS 123
OY 161 KLPEMLKMFEDRLSHKTYLNGDHTVHPDMLYDALDVLYMDPCLDAFPKLVSEKRIE 220
DB 124 GLPTLKMNSDFLGRHLYLTSVSHVDFMYEALDCIRYLAPOCLDEFPKLEKRSRIE 183
OY 221 AIPQIDKYLKSKYIAMPLOGMOAFFGGGDHP 253
DB 184 DLPKIKAYMESEKRIKMPPLNSHVSFSGGDHP 216

Search completed: May 20, 2003, 11:35:06
Job time : 28.557 secs

```

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 10:12:24 ; Search time 68.8588 Seconds

(without alignments)
2986.328 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357
Sequence: 1 MDWLRLFLMAAASINAA.....QAAACAPDLPAFSHGFSHN 998

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP-archaea:*
- 2: SP-bacteria:*
- 3: SP-fungi:*
- 4: SP-human:*
- 5: SP-invertebrate:*
- 6: SP-mammal:*
- 7: SP-mhc:*
- 8: SP-organelle:*
- 9: SP-phage:*
- 10: SP-plant:*
- 11: SP-rodent:*
- 12: SP-virus:*
- 13: SP-vertebrate:*
- 14: SP-unclassified:*
- 15: SP-virus:*
- 16: SP-bacteriap:*
- 17: SP-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3405	63.6	765	6	09TRK6
2	3370	62.9	765	11	09R055
3	1404.5	26.2	751	11	08VC36
4	1032	19.3	447	11	08R229
5	1029	19.2	218	5	094745
6	721	13.5	218	5	025595
7	707	13.2	218	5	09XYL9
8	649	12.1	787	3	096X16
9	615.5	11.5	271	11	0920K6
10	518.5	9.7	218	6	09NOV4
11	513.5	9.6	219	13	090VM9
12	507.5	9.5	218	11	08R516
13	503.5	9.4	223	6	097117
14	484.5	9.0	218	6	09TSM5
15	481.5	9.0	218	6	09TSM4
16	481.5	9.0	221	6	09BE50

17	479.5	9.0	218	11	091Y83	091Y83 cavia porce
18	472.5	8.8	204	11	09CW16	09CW16 mus musculu
19	468.5	8.7	218	11	09WU21	09WU21 rattus norv
20	464.5	8.7	218	11	09DC68	09DC68 mus musculu
21	463.5	8.7	218	11	09DD25	09DD25 mus musculu
22	463.5	8.7	219	5	027653	027653 echinococcu
23	462.5	8.6	219	5	016058	016058 echinococcu
24	455.5	8.5	225	4	096HA3	096HA3 homo sapien
25	449.5	8.4	225	6	09BEA9	09BEA9 macaca fusc
26	443.5	8.3	220	4	08WME1	08WME1 homo sapien
27	441.5	8.2	232	11	09D5T8	09D5T8 mus musculu
28	440.5	8.2	225	11	0921B2	0921B2 rattus norv
29	439.5	8.2	195	4	095465	095465 homo sapien
30	415.5	7.8	188	6	09M2B4	09M2B4 capra hircu
31	388	7.2	741	1	09C6V7	09C6V7 arabidopsis
32	386.5	7.2	219	5	09U582	09U582 psoraleps o
33	381	7.1	654	16	0989X2	0989X2 rhizobium l
34	372.5	7.0	668	10	048552	048552 arabidopsis
35	368.5	6.9	660	17	097XW1	097XW1 sulfolobus
36	358	6.7	64	4	09UEU7	09UEU7 homo sapien
37	357.5	6.7	756	10	09STI2	09STI2 arabidopsis
38	357	6.7	194	4	09HMT7	09HMT7 homo sapien
39	353.5	6.6	650	10	023349	023349 arabidopsis
40	353	6.6	649	10	08W1C1	08W1C1 brassica ju
41	350	6.5	181	4	08TC98	08TC98 homo sapien
42	339	6.3	735	10	09SM88	09SM88 canavalia l
43	338.5	6.3	1794	10	09S168	09S168 arabidopsis
44	335.5	6.3	422	10	09C6W1	09C6W1 arabidopsis
45	330.5	6.2	220	5	08T7E4	08T7E4 boophilus m

ALIGNMENTS

RESULT 1

ID 09TRK6 PRELIMINARY: PRT: 763 AA.

AC 09TRK6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Semicarbazide-sensitive amine oxidase (EC 1.4.3.6).
GN SAO.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovine; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Iwabuki H., Matsumura K., Mure M., Kuroda S., Tanizawa K.;
RT "Molecular cloning of semicarbazide-sensitive amine oxidase gene from
RT Bovine aorta.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB019242; BAA8896.1;
DR InterPro: IPR000269; CuniH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; UNKNOWN_1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
KW Oxidoreductase.
SQ SEQUENCE 763 AA; 84500 MW; DB61ED9A89E71E90 CRC64;

Query Match 63.6%; Score 3405; DB 6; Length 763;
Best Local Similarity 85.1%; Pred. No. 9.5e-267;
Matches 626; Conservative 44; Mismatches 66; Indels 0; Gaps 0;
QY 263 QGGGGGPPSOLPHCPSPSPSAQPTWHPGQSLFADLSREELTAVMRFTLQRLGGLVDA 322
Db 28 RGGGGGASOPHYCPSPSPSPSAQPTWHPGQSLFADLSREELTAVMRFTLQRLGGLVDA 87

QY	323	QARSMDCVSVELQLEDPKKAALAHLDKRGSPPARRELATIVFGORQOPVNSVSELYVGLP	382
QY	323	QARSMDCVSVELQLEDPKKAALAHLDKRGSPPARRELATIVFGORQOPVNSVSELYVGLP	382
Db	88	QARSDNCISVSEVLQLEPPKKAALAHLDKRSPPPARRELATIVFGOGQOPVNTELYVGLP	147
QY	383	HPXYMRDVTYERHAGCPPLPYHRBPVLPOEYLDIDOMINRELPOASGLHHCCCFYKHGRN	442
Db	148	QPSIMRDVYTERHAGCPPLPYRRPVLREYLDIDOMITNRELPOAAGVLHHCCSTYKOGGN	207
QY	443	LVTMTAPRGLSGDRATWGLYYNISAGFELHHVGLLELVNHHKALDPAKMTIOKVYQ	502
Db	208	LVTMTAPRGLSGDRATWGLYYNISAGYVLPVGLLELVNHHKALDPAQMTIOKVFEQ	267
QY	503	GRRYDLSLAQLEAOFEDALVNVVLIIPDNGTSGSNLSKPYRPPGAPPLQGFPOGPRFSVOG	562
Db	268	GRRYDLSLAQLEOFEADRVNVVLIIPNNGTSGSNLSKQVPPGPPPLQGFPOGPRFSVOG	327
QY	563	SRVASSLWTEFSFGIAGSCGRPIEDFVRQGERLVEISLOEALAIYGGNSPAAATTRVYDQ	622
Db	328	SRVASSLWTEFSFGIAGSGRPIEDIRQGERLAEISLOEVALIYGGNTPRAMLTRMDG	387
QY	623	GFGMGKTYTPLTRGVDCPYLATVYDMHFLLESQAPKTIIRDAFCVFEQNOGLPLRRHNSDL	682
Db	388	CFGMGKFTPLTRGVDCPYLATVYDMHFLLESQAPRTLDHAFCVFEQNKGLPLRRHNSDF	447
QY	683	YSHFPGGLAEVLVVRBSMSTLLANDVYDWDVYFHHSGAIEIRFYATGYISSFFELFGALGX	742
Db	448	ISOYFGGVVEVLVVRFSVSTLLANDVYDWDVYFHHNGAIEVKEFHAATGYISSFFELFGALGX	507
QY	743	GNQVSEHTLGTVHSHSAHFKVDLDVAGLENNVAAEDVAFVMAVPMSPEHOLQVYTRK	802
Db	508	GNQVRENTLGTVHSHSAHYKKVDLVGGLLENNVAAEDVAFVPTTYVMSPEHOLQVYTRK	567
QY	803	LLEMEQAAFLVGSATPRYILYLASNHSNKMGHPGRYRIOMLSFAGEBLLPONSMARGFSW	862
Db	568	QLETEQAAFLVGLGASPRYILYLASKSNKMGHPGRYRIQVYTSFAGRLLPONSTERALISW	627
QY	863	ERYQLATVQRKEEPPSSSVNQNDDPAAPTYDESDFINNETIAGKDLVAVYTAGFLIHP	922
Db	628	GRYQLATVQRKEETPSSSVNQNDDPPTPYDFADFINNETIAGKDLVAVYTAGFLIHP	687
QY	923	AEDIPNTVYVNGGFFLRAPNPFDEDEPSFASDSIYFRRGQDAGCEVNPILATLPAAA	982
Db	688	AEDIPNTVYVNGGFFLRAPNPFDEDEPSINSADSIYFOKKQDAGCEVNSLACLPRDPA	747
QY	983	CAPDLPAFASHGCGFSHN 998	
Db	748	CAPDLPAFASHGCGFTN 763	
RESULT 2			
O9R055 PRELIMINARY; PRT; 765 AA.			
AC	O9R055		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Copper amine oxidase (EC 1.4.3.6).		
GN	ACC3.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99194802; PubMed=10092636;		
RA	Moldes M., Feve B., Paltralt J.;		
RT	"Molecular cloning of a major mRNA species in murine 3T3 adipocyte		
RT	lineage, differentiation-dependent expression, regulation, and		
RT	identification as semicarbazide-sensitive amine oxidase.";		
RL	J. Biol. Chem. 274:9515-9523(1999).		
CC	-I- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +		
CC	H(2)O(2).		
CC	-I- COFACTOR: COPPER AND TOPAQUINONE (BY SIMILARITY).		

CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
DR EMBL; AF151411; AAD09189.1; -.
DR MGD; MG1:1306797; Aoc3.
DR InterPro; IPR000269; CuNH_oxidase.
DR Pfam; PF011179; Cu_amine_oxid; 1.
DR Pfam; PF02727; Cu_amine_oxidn; 1.
DR Pfam; PF02728; Cu_amine_oxidn; 1.
DR PRINTS; PR00766; CUDAOXIDASE.
DR PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE; PS01165; COPPER_AMINE_OXID_2; 1.
KW Copper: oxidoreductase: TPQ.
SQ SEQUENCE 765 AA; 84534 MW; 4252854759F316C7 CRC64;

Query Match	62.9%;	Score 3370;	DB 1;	Length 765;
Best Local Similarity	83.1%;	Pred. No. 6.5e-264;		
Matches 61;	Conservative 61;	Mismatches 63;	Indels 0;	Gaps 0;

QY	263	OGDGGESDLPJPHCSVSPASQPMWHPQSOJFALSLSBEELTAVNRBFLTORGLDYDA	322
Db	28	RSGDDGGJLSQPLHCSVSLPVSQPTNHPQSOFJFALSPEELTAVNSFLTKHGLPCLYDA	87
QY	323	QARPSDNCVFSVELQJPKAAALAHLDKRGPPPARREALATVEFGKQPOPNSVELVGP	382
Db	88	QARPSDNCVFSVELQJPKAAALAHLDKRGPPVPEEALAIFFGQKPPKNSVELVGP	147
QY	383	HSSVARDVTVEHNGRPLRVHREPVLEFOEJLDIDOMIFNRBELPQASGLLHCCFYHGRN	442
Db	148	HSSVARDVTVEHNGRPLRVHRRPVLDREJODILEEMIIFHELEQASGLLHCCFYHGOGN	207
QY	443	LVTMTTAPRGJOSGRATWFGLYNISGAGFPLJHNGFLELLAVNHNALDPAWTKQVYO	502
Db	208	LVTMTTAPRGJOSGRATWFGLYNLSGAGFYPHPIGIELLIDHKALDPAWTKQVYO	267
QY	503	GRTYSLSLAOLEQFENAGLVNVTLLIDNGTGGSMJLSKPPRPPRAPLPQFPGGPRFSVOG	562
Db	268	GRTYSLSLSLAOLEQFENAGLVNVTLLPNNGGSMJLSKSSPPRAPLPQFPGGPRFSVOG	327
QY	563	SKVASSLMTFESGJLAFSGPRIFDVRFQOGERLYVEISLOEALATYGNSPAAWTKRYDG	622
Db	328	SKVASSLMTFESGJLAFSGPRIFEDRIFQGERAYELSYOEALATLGGNSPAAWTKRYDG	387
QY	623	GFQMGKTYTPTLRGYVDCPYLATIYDWMHLLFSQAPKTIIDAFCVBEONQGLPLRRHSD	682
Db	388	GFQMGKTYTPTLRGYVDCPYLATIYDWMHLLFSQAPKTIIDAFCVBEONQGLPLRRHSD	447
QY	683	YSHYFGGLAEYLTLYVRSNMSTLNTDYVMDTYPHPSGAIEIRFYATGYISSAPLFGATGK	742
Db	448	YSHYFGGVGYTLTYVRSVSTLNTDYIMDMYHHPGALIEYKHNAGIYISSAPLFGATGK	507
QY	743	GNQVSEHTLGTVHTSHAEKVDLDVAGLENNWMAEDMVFVPAVWSPSEHOLQRLQVTK	802
Db	508	GNQVSAHTLGTVHTSHAEKVDLDVAAGLKNMAAEDMFAFVPIVWQPEYQKQRLQVTK	567
QY	803	LLEMEQOAFILVGSAPRRLYLASHNSNMKGPRGRIOMLSFAGEPLPONSMAKGSF	862
Db	568	LLEMEQOAFILVGSATPRLLYLASHNSNMKGRRGRIOLILFAGKPLPOESPIEKATF	627
QY	863	ERYOLAVTORKEEPESSSSVFQNDPMATVDSDFINNETIAGKDLVAMVTAGFLHP	922
Db	628	GRYHLAVTORKEEPESSSSIFQNDPMPTVYFTDFISNETIAGEDLVAMVTAGFLHP	687
QY	923	ARDINTVTVGNGVPEFLRPVNFPEDEPFSFYADSIFYPFGDDADAGEVNPJLACPOAA	982
Db	688	ARDIPTVTVAGNSVGFPLRPYNFDEDPFSHADSIFYREGDATRACEVNPJLACLSQAT	747
QY	983	CAPDLPAFSGHGFSH 997	
Db	748	CAPDLPAFSGHGFAY 762	
RESULT 3			
ID	Q8VC36	PRELIMINARY.	PRT. 751 AA.

AC 08VC36:
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Similar to amiloride binding protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-COLON;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC021880; AAH21880.1; -
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxid_2.
 DR Pfam: PF02728; Cu_amine_oxid_3.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; UNKNOWN_1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; UNKNOWN_1.
 SQ SEQUENCE 751 AA; 85449 MW; 178B430795A2B988 CRC64;

Query Match 26.2%; Score 1404.5; DB 11; Length 751;
 Best Local Similarity 39.8%; Pred. No. 1.1e-104;
 Matches 298; Conservative 127; Mismatches 270; Indels 53; Gaps 15;

QY 280 SPFAQWTHPGQSOLFADLSREELTAVMRELTORLPGVDAQAAPSDNCFVSEVETLP 339
 DB 19 TPEAVTTPH-DKARIFADLSPOEIKAVHSLMRKELGESSNLTAKSVLEIELLP 77
 QY 340 PKAALAHLDGRSPPRPRLAIVFPGROPVNSLVGRLPHPSYMRDVTYERHGGPL 399
 DB 78 KKNVNLKFLDEGRKSPVREARAIFFGAQDPVTEFAVGPRLPRPCVQALS-PRPGHNL 136
 QY 400 PYHRRVVLQOEYLDIDQMIFNRELPAQSGILHHCFCYKHHG-----NRLVTMTAPR 451
 DB 137 SMSRSPRISTAEY-DLTYHMLNRAITPL-----HOFELDTGFSFGCDDFLFTDYAPR 190
 QY 452 GLOSGDRATWFGLYUNISGAFGLHNVGELLYNHNKALDPAWRTIQKVFYGRGYDLSAQ 511
 DB 191 GVESGGRSMVLIQVRYE--GYFLHPTGLIIVDHSSTDVQDMVDEGLWNGKYNPEE 248
 QY 512 LEAFEGALVNVLLIPDNGTGSNSKSPVPG-PAPPL-----QFY----- 552
 DB 249 LAQKYAVEAEVAVLEE-----VLEEDPLPGATEQPRPFSSYKRGEFHTPVTVAGPHV 302
 QY 553 --PQRPFSVQGSRYVASLWTFSGLGAFSGPRFEDVRFQGERLVYEISLOEALATYGN 610
 DB 303 VQGPSPRYKLEGNNVLYGDMFSYRLRSSSGLOLFENVLFGGERVAYESVOEAVALTGGH 362
 QY 611 SPAAATRYVDGFGMKYKTTPLTRGVDCPYLATYVDMHFLLESQAQKTIIDAFCEVEON 670
 DB 363 TPAGMOKYITDVGWGLSVTHLAPGIDCEPETAFLDAFHYDSDGVLYPRALCLEMP 422
 QY 671 QGLPLRRHNSDLYS--HYFGGLAETVLYVRSKSTLLNDYVMDTVHPSCAIEIRYAT 727
 DB 423 TGVDLRHNFDSNFKGCFNYAGLKGVLVLTSTVYNYDYIMDFIYFVPCVMETKMAT 482
 QY 728 GYSSAFLFGATGKYGNVSEHTLGTVHTSHAHFKVDLDVAGLNNWMAEDMVFVMAVP 787
 DB 483 GYVHATYTPTEGRLHGRLOTHLGNHTHLVHYRVDLDVAGTNSFTLTCLKLENTINP 542
 QY 788 WSPHQLOLQVTRKLLMEEOAFLVGSATPRYLTLASNSKMGHPRGRYRIOMLSPAG 847
 DB 543 WSPHSLVQPTLEBOTYSHENQAAPFRGQTLPKYLTFSSPKKNMGHRSRYLOLHSMAE 602
 QY 848 EPLPONSMAFGSWERYQLAVTORKEEPPSSSVFNONDPWAPLVDSDFI-NNETIAG 906
 DB 603 QVLPQGOEERAVTWARPYLAIVTKYRESERYSSSLYNONDPWDPVVEEFLNNENIEN 662

QY 907 KDLVAVTAGEFLIHPAEDIPTVTVGNGVGFELREYNFEDDPSFYSDSIFYRQDODA 966
 DB 663 EDLVAVTVYGFELIHPSEDEVPNTATPAGNCVGFILRFNFEEEDPSLASDVTIV-WPQON 721
 QY 967 GACEVNPACLDPQAACAPDLPAFSGHG 994
 DB 722 GLNHVO--RWIPENRDCLYS-PPFSYNG 746

RESULT 4
 ID Q8R229 PRELIMINARY; PRT; 447 AA.
 AC Q8R229;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical 51.3 kDa protein (fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-COLON;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC022627; AAH22627.1; -
 KW Hypothetical protein.
 FT NON-TER
 SQ SEQUENCE 447 AA; 51265 MW; E556E4AA2525D1D1 CRC64;

Query Match 19.3%; Score 1032; DB 11; Length 447;
 Best Local Similarity 44.6%; Pred. No. 7e-75;
 Matches 199; Conservative 78; Mismatches 161; Indels 8; Gaps 5;

QY 553 PQGPRESVQGSRYVASLWTFSGLGAFSGPRFEDVRFQGERLVYEISLOEALATYGNP 612
 DB 1 PGGPRYKLEGNNVLYGDMFSYRLRSSSGLOLFENVLFGGERVAYESVOEAVALTGGHTP 60
 QY 613 AAMTRVVDGFGMGKXTTPLTRGVDCPYLATYVDMHFLLESQAQKTIIDAFCEVEONOG 672
 DB 61 AGMOTKTIIDVGWGLSVTHLAPGIDCEPETAFLDAFHYDSDGVLYPRALCLEMPGTG 120
 QY 673 LPLRRHNSDLYS--HYFGGLAETVLYVRSKSTLLNDYVMDTVHPHSGAIEIRFYATGY 729
 DB 121 VPLRRHFDNSNKGCFNYAGLKGVLVLTSTVYNYDYIMDFIYFVPCVMETKMATGY 180
 QY 730 ISSAFLEGCATGKYGNVSEHTLGTVHTSHAHFKVDLDVAGLNNWMAEDMVFVMAVPMS 789
 DB 181 VHAFTYTPTEGRLHGRLOTHLGNHTHLVHYRVDLDVAGTNSFTLTCLKLENTINPMS 240
 QY 790 PEHLOLOLQVTRKLLMEEOAFLVGSATPRYLTLASNSKMGHPRGRYRIOMLSPAGEP 849
 DB 241 PSHSLVQPTLEBOTYSHENQAAPFRGQTLPRYLTFSSPKKNMGHRSRYLOLHSMAEQV 300
 QY 850 LPONSSMAFGSWERYQLAVTORKEEPPSSSVFNONDPWAPLVDSDFI-NNETIAGKD 908
 DB 301 LPQGOEERAVTWARPYLAIVTKYRESERYSSSLYNONDPWDPVVEEFLNNENIENED 360
 QY 909 LVAVTAGEFLIHPAEDIPTVTVGNGVGFELRPNFEDDPSFYSDSIFYRQDODGA 968
 DB 361 LVAVTAGEFLIHPSEDEVPNTATPAGNCVGFILRFNFEEEDPSLASDVTIV-WPDNGL 419
 QY 969 CEVNPACLDPQAACAPDLPAFSGHG 994
 DB 420 NHVO--RWIPENRDCLYS-PPFSYNG 442

RESULT 5
 ID Q94745 PRELIMINARY; PRT; 218 AA.
 AC Q94745;
 DT 01-FEB-1997 (Tremblrel. 02, Created)

DT 01-FEB-1997 (TREMBlrel. 02, last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
 DE 26kd glutathione S-transferase.
 OS Schistosoma mekongi.
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=38744;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=THAILAND.
 RX MEDLINE=9822295; Pubmed=9561610;
 RA Grams S.V., Grams R., Korge G., Vyanant V., Upatham S.;
 RT "Cloning and sequencing of the 26 kDa glutathione S-transferase gene
 of Schistosoma mekongi."
 RL Southeast Asian J. Trop. Med. Public Health 28:570-574(1997).
 DR EMBL: Y07663; CA68944.1; -.
 DR HSSP: P08515; 1GTA.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase.
 SQ SEQUENCE 218 AA; 25486 MW; D3080620B19DE23A CRC64;
 Query Match 19.2%; Score 1029; DB 5; Length 218;
 Best Local Similarity 88.9%; Pred. No. 3.9e-75;
 Matches 193; Conservative 11; Mismatches 13; Indels 0; Gaps 0;
 OY 38 SPILGYWKIKGLVOPTRLLLEYLEKEEYEHLYERDEGDKWRKKFELGLEFPNLPYYIDG 97
 DB 2 APLVGYWKIKGLVOPTRLLLEYLEGEYERLYERNEGDAMREKFKLGEFPNLPYYIDG 61
 OY 98 DVKLQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRVSKDFETLKVD 157
 DB 62 DVKLQSAIIRYIADKHNMLGSGPKERAISMLGAVSDIRSGVSRVSKDFETLKVD 121
 OY 158 FLSKLPKMKFEDRLSHKTYLNGDHYTHPDPMYDALDVLYMDPCIDAFPKLVSEFK 217
 DB 122 FLNKLPEMLKMFEDRLCHKTYLNGDRVTHPPFMYDALDVLYMDPCIDAFPKLVSEFK 181
 OY 218 RIEAIPQIDKYLKSSKYIAMPLOGQATFGGDDHPK 254
 DB 182 RIEAIPQIDKYLKSSKYIAMPLOGQATFGGDDHPK 218
 RESULT 6
 ID 025595 PRELIMINARY: PRT: 218 AA.
 AC 025595;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative glutathione transferase.
 OS Clonorchis sinensis.
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Opisthorchiida; Opisthorchiata; Opisthorchiidae; Opisthorchiidae;
 OC Clonorchis.
 OX NCBI_TaxID=79923;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hong S.-J., Lee D.-H.;
 RT "Cloning and over expression of 26 kDa glutathione S-transferase from
 Clonorchis sinensis."
 RL Thesis (1996), Parasitology, Chung-Ang University College of Medicine,
 Seoul.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hong S.-J.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: L47992; AAB46369.3; -.
 DR HSSP: P31670; 1FHE.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.

DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase.
 SQ SEQUENCE 218 AA; 25038 MW; 7CB17C7B837A0B7C CRC64;
 Query Match 13.5%; Score 721; DB 5; Length 218;
 Best Local Similarity 59.0%; Pred. No. 3.5e-50;
 Matches 128; Conservative 38; Mismatches 51; Indels 0; Gaps 0;
 OY 38 SPILGYWKIKGLVOPTRLLLEYLEKEEYEHLYERDEGDKWRKKFELGLEFPNLPYYIDG 97
 DB 2 APLVGYWKIKGLVOPTRLLLEYVGSYEHSHYRCDEGDKWRKKFELGLEFPNLPYYIDG 61
 OY 98 DVKLQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRVSKDFETLKVD 157
 DB 62 NFSLQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRVSKDFETLKVD 121
 OY 158 FLSKLPKMKFEDRLSHKTYLNGDHYTHPDPMYDALDVLYMDPCIDAFPKLVSEFK 217
 DB 122 YLQQLPSTLRWMSQFLGNNYSYLGSTPHLDPMFYEALDVIRYDPTSGEAFPNLMQFIH 181
 OY 218 RIEAIPQIDKYLKSSKYIAMPLOGQATFGGDDHPK 254
 DB 182 RIEAIPQIDKYLKSSKYIAMPLOGQATFGGDDHPK 218
 RESULT 7
 ID 09XYL9 PRELIMINARY: PRT: 218 AA.
 AC 09XYL9;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Glutathione S-transferase.
 GN GST-1.
 OS Fasciola gigantica.
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Schistosomatidae; Echinostomata; Fasciolidae; Fasciolidae; Fasciola.
 OX NCBI_TaxID=46835;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=THAILAND.
 RA Grams S.V., Grams R., Sobhon P., Vyanant V., Upatham S.;
 RT "Molecular cloning of expressed antigens from Fasciola gigantica."
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF112567; AAD23997.1; -.
 DR HSSP: P31670; 1FHE.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase.
 SQ SEQUENCE 218 AA; 25337 MW; E9423D75C3F2EAF CRC64;
 Query Match 13.2%; Score 707; DB 5; Length 218;
 Best Local Similarity 58.7%; Pred. No. 4.7e-49;
 Matches 125; Conservative 36; Mismatches 52; Indels 0; Gaps 0;
 OY 41 LGYWKIKGLVOPTRLLLEYLEKEEYEHLYERDEGDKWRKKFELGLEFPNLPYYIDG 100
 DB 5 LGYWKIKGLVOPTRLLLEYLEGEYERLYERNEGDAMREKFKLGEFPNLPYYIDG 64
 OY 101 LTQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRVSKDFETLKVD 160
 DB 65 LTQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRVSKDFETLKVD 124
 OY 161 KLPKMKFEDRLSHKTYLNGDHYTHPDPMYDALDVLYMDPCIDAFPKLVSEFK 220
 DB 125 ELPTLKMSQFLGNNYSYLGSTPHLDPMFYEALDVIRYDPTSGEAFPNLMQFIH 184
 OY 221 AIPQIDKYLKSSKYIAMPLOGQATFGGDDHPK 253
 DB 185 DLPKIKAYMESEKTIKMPINLSWTASFGGDDAP 217

RESULT 8

096x16 PRELIMINARY; PRT; 787 AA.

ID 096x16
AC 096x16
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Lysyl oxidase.
GN AOC1.
OS Pichia pastoris (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4922;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21131996; PubMed=11237259;
RA Kucha J.A., Dooley D.M.,
RT "Cloning, sequence analysis, and characterization of the 'lysyl
RT oxidase' from Pichia pastoris."
RL J. Inorg. Biochem. 83:193-204(2001).
DR EMBL: AF358434; AAK49976.1; -
DR Interpro: IPR000269; Cunn.oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
SQ SEQUENCE 787 AA; 89650 MW; D779E8F2BF766745 CRC64;

Query Match 12.1%; Score 649; DB 3; Length 787;
Best local similarity 27.3%; Pred. No. 1.9e-43;

Matches 201; Conservative 114; Mismatches 320; Indels 102; Gaps 21;

OY 280 SPQAQWTHGQSLRADLSREELTVAMRFLTRLQRLGVLDAQAQPSDNCVSVELQPL 339
DB 55 APTNTNT-----SLAKEVQVLDLHSTY--NITEVTKAFVSVYVLMIEFLK 103
OY 340 PKAALAHLDGSPPREALALVFGROPVSELVLPHPSPMYRVTERRHGRL 399
DB 104 NKTEALTYLDEDDLPFRNARTVYFGEGEGFEELKVPPL---VSDETITE---PL 156
OY 400 PYHRRVFLQEVYLDIDQMTFNRLPOASGLLHCCEYKARGRLVTTTPARGL----- 453
DB 157 SFNTNGKSKLPREVGLHRIKSAKSS-----FLKNLNTTIRVDLHGLIGVPE 208
OY 454 -----QSGRATWFGLYT-----NISGAFELHVLVLELVNKAIDPA 492
DB 209 DMGCHSAAPQLHDPATGATVDTGTCNINTENDAEMLVPTGFEFR-----FDTGGRDVS 261
OY 493 RMTIOKVFYGGRYVDSLAOL-BAOFEAGLVNVVLI-P-DNGTGSMSLSKSPVPPGPR----- 545
DB 262 QMKLEYITNNKYTSAEELYEAMQKDDFTLKRIDVN---LDMTYIQORDSAPRINHL 318
OY 546 --APPLQFYQGRFVSQGSRVASL--WTFSEGLGAFSGPRIFDVAFQGERLYEISLQ 601
DB 319 DRSPRLVEDEGRMAVDGEEYFSMDMGVYTSMSDGTISFDITFKSGRIYELSLQ 378
OY 602 EALAIIGGSPRAATTYYVDGEGMGKTTPLTRGVDCPIATATVDMH-FLLESQAQRTI 660
DB 379 ELIAEYSGDDPFNOHFTYSISYGVGNRFS-LVPGYDCPATAGYFTTDEFEYDEFYRNL 437
OY 661 RDAFCVEQNOGLPLRRHSDLSHYFGSLAETVLVRSMTLNTLVYDMVTFVHPSGAI 720
DB 438 --STCVENQEDVSLNHTGASYSAT---TONPLNVRISTIGNYDINFLYKFFLDGTL 492
OY 721 EIREYATGYISSAFLFGAT-GKYGNOVSEHTLGTVHTSAHKYKVDVAGLENNVMAEDM 779
DB 493 EVSVRAAGYIOAGWNPETSAFYGLKIHVLSGSFHDVAVLVKYDLDVCGTKNNASKVM 552
OY 780 VFVMAVWSPHQLQVTRKLEME-----QAFLVGSATPRYKYLASHS 829
DB 553 KDVOVEYAWAGTYVNTQIAREVLEKEDFNGINWPGQGIILIESA-----EET 603
OY 830 NKMGRPRGRTOMLSFAGEPLPONS-SMARGFSMERYOLAVTORKEEPPSSSVFNODP 888

DB 604 NSFGNPRAYNIMPGGCVHRIYKNSRGPETQNMARSNLFLTKHKDEELRSSTALNTNAL 663
OY 889 WAPTYDPSDFINNETIACKDLVAMYTACGLHTPHADIPNTYVNGVGFELRPYNFDE 948
DB 664 YDPVNFNAFLDESDLEDGDIYAVNVLGLHLPLNSNDLNTLTSTAHASFMLTPRYFDS 723
OY 949 DPFYSADSIYFRGDD 965
DB 724 ENSRDTQOVFTTYDDE 740

RESULT 9

0920K6 PRELIMINARY; PRT; 271 AA.

ID 0920K6
AC 0920K6
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Diamine oxidase (EC 1.4.3.6) (Fragment).
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognath; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=SMALL INTESTINE;
RA Kitanaka J., Kitahara N., Takemura M.,
RT "Cloning of a cDNA for guinea-pig diamine oxidase (histaminase)."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBD databases.
DR EMBL: AB073298; BAB70498.1; -
DR Interpro: IPR000269; Cunn.oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; UNKNOWN_1.
KT Oxidoreductase.
FT NON_TER
SQ SEQUENCE 271 AA; 31323 MW; CEC46D4EAA135D CRC64;

Query Match 11.5%; Score 615.5; DB 11; Length 271;
Best local similarity 49.1%; Pred. No. 1.7e-41;

Matches 115; Conservative 31; Mismatches 87; Indels 1; Gaps 1;

OY 726 ATGYISSAFLFGATGKYGNOVSEHTLGTVHTSAHKYKVDVAGLENNVMAEDVPEYVNA 785
DB 1 ATGVNATFTYTPGQRTGLHTLGNHTLHLYVVDVAGTKKNSFOTLQKLENT 60
OY 786 VPMSPHQRLQVTRKLEMEQDAFLVSATPRYKYLASHSNNKKGHRGRYQVLSF 845
DB 61 NPMSPGHHLVQPTLKOSRYSOERQAFRFQOTLPKYLFTSREKNNRNGHGRYQVLSF 120
OY 846 AGEFLPNSSMARGFMSERYOLAVTORKEEPPSSSVFNODPAPTYDPSDFI-NNETI 904
DB 121 AEQVLPAGLDEPERAITVARIPLAVTKYRESELCSSYIYNDPDPVYVOKFLHNEDI 180
OY 905 AGKDLVAMVTAGFLPHIADIPNTVTVNGVGFELRPYNFDEDPSPYSADSI 958
DB 181 ENEDLVAMVTYGFILHHSDEVPYATPAGNSVGLPLRPNFNFKDPSLASHDVT 244

RESULT 10

09NOV4 PRELIMINARY; PRT; 218 AA.

ID 09NOV4
AC 09NOV4
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Class mu glutathione S-transferase.
GN GSTM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

RA MEDLINE-99381232; PubMed-10451925; G.W., George J.E.;
 RA He H., Chen A.C., Davey R.B., Ivie G.W., George J.E.;
 RT "Characterization and molecular cloning of a glutathione S-transferase
 RT gene from the tick, *Boophilus microplus* (Acari: Ixodidae).";
 RL Insect Biochem. Mol. Biol. 29:737-743(1999).
 DR EMBL: AF077609; AAD15991.1; -
 DR HSSP: P20136; 1GSU.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 DR Transferrase.
 KW SEQUENCE 223 AA; 25604 MW; F276ACE2D8259174 CRC64;

Query Match 9.4%; Score 503.5; DB 5; Length 223;
 Best Local Similarity 47.0%; Pred. No. 1.5e-32;
 Matches 101; Conservative 31; Mismatches 78; Indels 5; Gaps 1;

OY 38 SPILGYWKIKGLVPTRLLEYLEEKYEHLERD-----EGDKRNKKFELGLEFPNLP 92
 DB 2 APLVGYMDIRGLAOPIRLLAHADAKYDKRYCGPPEDRSSWLNKTKLGLFPNLP 61
 OY 93 YIIDGVKLTQSMATIRYIAOKHNMGGSPKRAEISMLEGAVLDIRYGVSRIVASKDFE 152
 DB 62 YIIDGVKLTQSMATIRYIAOKHNMGGSPKRAEISMLEGAVLDIRYGVSRIVASKDFE 121
 OY 153 TLKVDPLSKLPEMLKMFEDRLSHKTYLNGDHTVHPDPMYDALDVLYMDPCLDAFPKLV 212
 DB 122 KTKGDIKLNPLAKAFSDYLGTHKRFAGDNLTLYVDFIAEMLAQHILFAPDCLKPFLNL 181
 OY 213 VSEFKRIEAIPOIDKYKSSKYIAMPLOGWQATFG 247
 DB 182 KAFVDRIEALPHVAAYLKSDKICWPLNGDMASFG 216

RESULT 14
 OYTSMS PRELIMINARY; PRT; 218 AA.
 AC OYTSMS;
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase mu-class subunit M1 (EC 2.5.1.18).
 GN GSTM1.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE-20330602; PubMed-10869451;
 RA Wang C., Bammeler T.K., Guo Y., Kelly E.J., Eaton D.L.;
 RT "Mu-class GSTs are responsible for aflatoxin B(1)-8, 9-epoxide-
 RT conjugating activity in the nonhuman primate macaca fascicularis
 RT liver.";
 RL Toxicol. Sci. 56:26-36(2000).
 DR EMBL: AF200709; AAF08539.1; -
 DR HSSP: P09488; 1GTU.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 DR Transferrase.
 KW SEQUENCE 218 AA; 25577 MW; EE5B4D80F60C95EB CRC64;

Query Match 9.0%; Score 484.5; DB 6; Length 218;
 Best Local Similarity 43.9%; Pred. No. 5e-31;

Matches 93; Conservative 37; Mismatches 77; Indels 5; Gaps 1;
 OY 41 LGYWKIKGLVPTRLLEYLEEKYEHLERDEG-----DKRNKKFELGLEFPNLPYI 95
 DB 5 LGYWKIRGLAHAIRLLETTDSSYEKKYTMGDAPDYDRSQWLNKFKGLDFPMLPYLI 64
 OY 96 DGDVLTQSMATIRYIAOKHNMGGSPKRAEISMLEGAVLDIRYGVSRIVASKDFETLK 155
 DB 65 DGTHTKITSNATIRYIAOKHNMGGSPKRAEISMLEGAVLDIRYGVSRIVASKDFETLK 124
 OY 156 VDFLSKLPMLKMFEDRLSHKTYLNGDHTVHPDPMYDALDVLYMDPCLDAFPKLVSF 215
 DB 125 PEYLEGPLPTMQHFSQFLKRPVEVDKITEVDLAYVDLHRIEPRCLDAFPMLKDF 184
 OY 216 KRIEAIPOIDKYKSSKYIAMPLOGWQATFG 247
 DB 185 ISHFEGLEKISAYMKSSRFLPRLYTRVAVWG 216

RESULT 15
 OYTSMS PRELIMINARY; PRT; 218 AA.
 AC OYTSMS;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase mu-class subunit M2 (EC 2.5.1.18).
 GN GSTM2.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE-20330602; PubMed-10869451;
 RA Wang C., Bammeler T.K., Guo Y., Kelly E.J., Eaton D.L.;
 RT "Mu-class GSTs are responsible for aflatoxin B(1)-8, 9-epoxide-
 RT conjugating activity in the nonhuman primate macaca fascicularis
 RT liver.";
 RL Toxicol. Sci. 56:26-36(2000).
 DR EMBL: AF200710; AAF08540.1; -
 DR HSSP: P28161; 1HNA.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 DR Transferrase.
 KW SEQUENCE 218 AA; 25708 MW; F2E509C3949F9051 CRC64;

Query Match 9.0%; Score 481.5; DB 6; Length 218;
 Best Local Similarity 44.8%; Pred. No. 8.7e-31;
 Matches 95; Conservative 35; Mismatches 77; Indels 5; Gaps 1;

OY 41 LGYWKIKGLVPTRLLEYLEEKYEHLERDEG-----DKRNKKFELGLEFPNLPYI 95
 DB 5 LGYWKIRGLAHAIRLLETTDSSYEKKYTMGDAPDYDRSQWLNKFKGLDFPMLPYLI 64
 OY 96 DGDVLTQSMATIRYIAOKHNMGGSPKRAEISMLEGAVDIRYGVSRIVASKDFETLK 155
 DB 65 DGTHTKITSNATIRYIAOKHNMGGSPKRAEISMLEGAVDIRYGVSRIVASKDFETLK 124
 OY 156 VDFLSKLPMLKMFEDRLSHKTYLNGDHTVHPDPMYDALDVLYMDPCLDAFPKLVSF 215
 DB 125 PEYLEGPLPTMQHFSQFLKRPVEVDKITEVDLAYVDLHRIEPRCLDAFPMLKDF 184
 OY 216 KRIEAIPOIDKYKSSKYIAMPLOGWQATFG 247
 DB 185 ISFEGLEKISAYMKSSRFLPRLYTRVAVWG 216

Search completed: May 20, 2003, 11:37:53
Job time : 72.8588 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:39:34 ; Search time 41 Seconds

(without alignments)
2418.011 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763

Perfect score: 4010
Sequence: 1 LVCVLVGRGGDGGEPGSLP.....QAAACAPDLPAFSGHGFSSHN 744

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

_A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4010	100.0	763	20	AAV03219
2	2489.5	62.1	729	23	AAU84261
3	457.5	11.4	248	23	ABP41516
4	329	8.2	712	23	ABB91435
5	324	8.1	687	23	ABB92504
6	302.5	7.5	759	23	ABB92082
7	246.5	6.1	300	23	ABB92910
8	214.5	5.3	670	17	AA94370
9	127	3.2	863	21	AA842952
10	125.5	3.1	460	23	ABB92909

11	115	2.9	2675	21	AA807564	Protein encoded by
12	109	2.7	435	14	AA832999	Rat choline kinase
13	106.5	2.7	446	22	ABG13153	Novel human diageno
14	106	2.6	617	23	AAE23407	Lysine oxidase pro
15	106	2.6	825	22	ABB67526	Drosophila melanog
16	104.5	2.6	594	22	ABG07023	Novel human diageno
17	104	2.6	1250	22	AA65914	Novel human diageno
18	103	2.6	2224	17	AAW04254	Novel human diageno
19	103	2.6	2224	17	AAW04254	Novel human diageno
20	102.5	2.6	541	20	AAV01648	Cytochrome P450 de
21	102.5	2.6	541	20	AAV01647	Cytochrome P450 de
22	102	2.5	1194	22	AA692078	C glutamylcum prote
23	101.5	2.5	587	22	ABB52856	Escherichia coli p
24	101.5	2.5	736	22	ABB57329	Escherichia coli p
25	101	2.5	617	23	AAE23403	Lysine oxidase pro
26	100.5	2.5	718	22	AAU54939	Protonibacterium
27	100.5	2.5	1253	23	AAU10648	Human L1CAM proteol
28	100.5	2.5	1257	20	AAW74152	Human L1 cell adhe
29	100.5	2.5	1544	23	AAU97541	Human phospholiposi
30	100	2.5	617	22	AAU02204	Escherichia coli p
31	98.5	2.5	844	22	ABB53083	Human secreted pro
32	98.5	2.5	897	22	AAU93422	Human secreted pro
33	98.5	2.5	933	20	AAV26349	Human secreted pro
34	98.5	2.5	933	21	AAV73400	Human secreted pro
35	98.5	2.5	1152	21	AAV94920	Human secreted pro
36	98.5	2.5	1494	23	AAU78460	Novel human diageno
37	98	2.4	810	22	ABG04698	Novel human diageno
38	98	2.4	929	20	AAV22189	Trimeric murine C3
39	97.5	2.4	308	20	AAV30811	Human secreted pro
40	97.5	2.4	903	21	AA842926	Human ORF25690
41	97	2.4	545	23	ABB91611	Herbicidally activ
42	96.5	2.4	763	22	AAU41222	Protonibacterium
43	96.5	2.4	816	23	ABB89783	Human polypeptide
44	96.5	2.4	1137	8	AAV71182	Sequence of herpes
45	96	2.4	1982	22	ABG16404	Novel human diageno

ALIGNMENTS

RESULT 1	AAV03219	standard: Protein: 763 AA.
XX	AAV03219;	
AC	21-JUN-1999 (first entry)	
XX		
DT		
XX		
DE	Amino acid sequence of the vascular adhesion protein-1.	
XX		
KW	Human; vascular adhesion protein-1; VAP-1; endothelial cell;	
KW	Lymphocyte; inhibition; amine oxidase.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 43	/note= "potential O-glycosylation site"
FT	Misc-difference 47	/note= "potential O-glycosylation site"
FT	Misc-difference 679	/note= "potential O-glycosylation site"
FT	Misc-difference 679	/note= "potential O-glycosylation site"
FT	Misc-difference 137	/note= "potential N-glycosylation site"
FT	Misc-difference 137	/note= "potential N-glycosylation site"
FT	Misc-difference 232	/note= "potential N-glycosylation site"
FT	Misc-difference 294	/note= "potential N-glycosylation site"
FT	Misc-difference 592	/note= "potential N-glycosylation site"
FT	Misc-difference 618	/note= "potential N-glycosylation site"
FT	Misc-difference 666	/note= "potential N-glycosylation site"

Query Match	Best Local Similarity	Score	DB	Length	Matches	744; Conservative	0; Mismatches	0; Indels	0; Gaps
1	100.0%	4010;	DB 20;	Length 763;	Best Local Similarity 100.0%; Pred. No. 0;	Matches 744; Conservative	0; Mismatches	0; Indels	0; Gaps
20	100.0%	4010;	DB 20;	Length 763;	Best Local Similarity 100.0%; Pred. No. 0;	Matches 744; Conservative	0; Mismatches	0; Indels	0; Gaps
61	100.0%	4010;	DB 20;	Length 763;	Best Local Similarity 100.0%; Pred. No. 0;	Matches 744; Conservative	0; Mismatches	0; Indels	0; Gaps
80	100.0%	4010;	DB 20;	Length 763;	Best Local Similarity 100.0%; Pred. No. 0;	Matches 744; Conservative	0; Mismatches	0; Indels	0; Gaps
121	100.0%	4010;	DB 20;	Length 763;	Best Local Similarity 100.0%; Pred. No. 0;	Matches 744; Conservative	0; Mismatches	0; Indels	0; Gaps

Db	140	ELVYGPLRPHRYMDVYVERIGGGLPRHRRVPLFQEVLDDIDQMIFFNELRPOASGLLHNC	199
Qy	191	FYKHGRNLVTMTAPRGLSGDRATVFGILYYNISGAFGLHHVGETLLVNHKALDPAR	240
Db	200	FYKHGRNLVTMTAPRGLSGDRATVFGILYYNISGAFGLHHVGETLLVNHKALDPAR	259
Qy	241	TIQKFTYGRYDLSLAOLEAFEGVLVNVLLIPNNGSGMSLKSPPRPPRAPPLQFPQ	300
Db	260	TIQKFTYGRYDLSLAOLEAFEGVLVNVLLIPNNGSGMSLKSPPRPPRAPPLQFPQ	319
Qy	301	GPRFVQGSRAVSSIMTFSFGLGAFSGPRIDVFFQGEERLYEISLQELALITGNSPAA	360
Db	320	GPRFVQGSRAVSSIMTFSFGLGAFSGPRIDVFFQGEERLYEISLQELALITGNSPAA	379
Qy	361	MTRRYVDGFGMGKXTTPPLTRGVDCPYLATVYDWHFLLSQAPRTIRDAFCVEQNGLP	420
Db	380	MTRRYVDGFGMGKXTTPPLTRGVDCPYLATVYDWHFLLSQAPRTIRDAFCVEQNGLP	439
Qy	421	LRHHSDLSYHFGGLATVLYVMSKSTLWYDVMPTVHFPSAIEIRRYATGYISSAF	480
Db	440	LRHHSDLSYHFGGLATVLYVMSKSTLWYDVMPTVHFPSAIEIRRYATGYISSAF	499
Qy	481	LEGATGKYGNOVSEHTLTSTHSAHFVDDVAGLENNVAAEDMVFVPAVAPVSPHQ	540
Db	500	LEGATGKYGNOVSEHTLTSTHSAHFVDDVAGLENNVAAEDMVFVPAVAPVSPHQ	559
Qy	541	ORLQYTRKLLKMEQAAFLVGSATPRYLXYLASNHSNKGHPRGYRIQLSPAGEPLPONS	600
Db	560	ORLQYTRKLLKMEQAAFLVGSATPRYLXYLASNHSNKGHPRGYRIQLSPAGEPLPONS	619
Qy	601	SWARFSEWRIOLAYTORKEEPPSSSYFNONDPAAPVDSFTINNETIAGKDLVAVT	660
Db	620	SWARFSEWRIOLAYTORKEEPPSSSYFNONDPAAPVDSFTINNETIAGKDLVAVT	679
Qy	661	AEFLIHIAHEDIPIWTVTGNNGVGFLLRPYNFFDEDPSTYSADSITYFGDDAGACEVNP	720
Db	680	AEFLIHIAHEDIPIWTVTGNNGVGFLLRPYNFFDEDPSTYSADSITYFGDDAGACEVNP	739
Qy	721	ACLPQAAACAPDLPAFSGHGSFN 744	
Db	740	ACLPQAAACAPDLPAFSGHGSFN 763	
RESULT 2			
AA084261			
ID	AA084261 standard; Protein; 729 AA.		
XX	AA084261;		
AC			
DT	08-MAY-2002 (first entry)		
XX			
DE	Human endometrial cancer related protein, AOC2.		
XX			
KW	Human; endometrial cancer; differential expression;		
XX	DNA microarray; protein microarray.		
OS	Homo sapiens.		
PN	WO200209573-A2.		
XX			
PD	07-FEB-2002.		
XX			
PF	31-JUL-2001; 2001WO-US24104.		
XX			
PR	31-JUL-2000; 2000US-221735P.		
XX			
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.		
XX			
PI	Mutter GL;		
XX			
DR	WPI: 2002-179967/23.		
XX	DR N-PSDB: ABR35481.		

PT Diagnosing endometrial cancer comprises determining expression of
PT nucleic acid molecules or expression products that are differentially
PT expressed in normal and malignant endometrium -
PS Claim 33; Page 131-134; 233pp; English.
XX
CC The invention relates to diagnosing endometrial cancer in a subject
CC suspected of having endometrial cancer comprising determining the
CC expression of a set of nucleic acid molecules or expression products in
CC an endometrial sample suspected of being cancerous, where the set of
CC nucleic acid molecules comprises at least 2 nucleic acid molecules
CC selected from 50 fully defined sequences as given in the specification.
CC The nucleic acids are used as an array of at least 2 of the 50
CC nucleic acids bound to a solid substrate. Also included is a solid-phase
CC protein microarray comprising at least 2 antibodies or its antigen
CC binding fragments, that specifically bind at least 2 different
CC polypeptides from the 50 fully defined sequences as given in the
CC specification, fixed to a solid substrate. The methods and arrays are
CC useful for the diagnosis of endometrial cancer, selecting and monitoring
CC treatment regimes and identification of lead compounds useful for the
CC treatment of endometrial cancer. The present sequence is one of 50
CC proteins differentially expressed between cancerous and non-cancerous
CC samples.
CC
CC
SQ Sequence 729 AA:
Query Match 62.1%; Score 2489.5; DB 23; Length 729;
Best Local Similarity 63.7%; Pred. No. 2,1e-244;
Matches 473; Conservative 86; Mismatches 149; Indels 35; Gaps 4;
QY 1 LVCVLLVGGGCGGEPSPHPCPSVSPSNOPTWPCQSOLFADLSREELTAVRFLTORL 60
DB 20 LAYVLLTSPGS---SQPHCPSPVSHRAQPMWPHQSQLFDLSREELTAVRFLTORL 75
QY 61 GGLVDAQAARSPDNCVFSEVLELPKKAALAHLDGRGSPPPAREALATYFFGQPOPNTS 120
DB 76 GGLVDAQAARSPDNCIFSEVLELPKKAALAHLDGRGSPPPAREALATYFFGQPOPNTS 135
QY 121 ELVVGELPHPSYMDVTVRHHGGLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHCC 180
DB 136 ELVVGELPHPSYMDVTVRHHGGLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHCC 194
QY 181 FKHGRNRLVMTTARGLQSGDRATWPLTYNISGAGFLHNVGELLVNHKALDPARM 240
DB 195 -FNNGSTLAAVAHAPRGARSRRRTTWIGYHNISGVGLFHLRVGLLELLDHALDPAHM 253
QY 241 TIOKVFYQGRYYDSLQLEAFGLVNVLLIPDNGTGSWSLSPVPGPAPPLQFYPO 300
DB 254 TYQOVYLLGHYADLQGLREFRSGLRYVPLPPNCASSLRKNSGQPLPLQFSPQ 313
QY 301 GPRFSYQGRVASSLWTFESFGAGSPRIFDVRFQGERLYVEISLQALATYGGNSPPA 360
DB 314 GSOYSYQGRVASSLWTFESFGAGSPRIFDVRFQGERLYVEISLQALATYGGNSPPA 373
QY 361 MTRRYVGGEGGKTYTPTLRGVDCPYLATYVDMHFLLESQAOKTIRDAFCVEQNGPL 420
DB 374 MLTRVLDSSFGGLRNSRGVLRGVDCPYQATMVDIHLVNGAQLLPGACVCEEQNGPL 433
QY 421 LRRHSHDLTSHYFGLAETVLLVRSMTLNDYVMDTVFHPGSALEIRPATYGSASF 480
DB 434 LRRHSHDLTSHYFGLAETVLLVRSMTLNDYVMDTVFHPGSALEIRPATYGSASF 493
QY 481 LFGATG--KYGNQVSEHTLGTVHTSHAHKRVLDLVAGLENMVAEDMVEPVMAWPSPH 538
DB 494 LKGGEGLLFGNRYGVRVLTGVTHTAHFHEKLDLDVAGLKMVVAEDVFEKPAVAPNPH 553
QY 539 QLORLQVTRKLLMEQAAFLVGSATPRYLYLASNSNKGHRGRIGRLQSLRAGEPLQ 598
DB 554 WLQRLQVTRKLLMEQAAFLVGSATPRYLYLASNSNKGHRGRIGRLQSLRAGEPLQ 598
QY 599 NSSMAGFSEWERYQLAOTQRKEEPPSSSVFNQNDPAPTPVDSFINNETIAGKDLVAN 658
DB 599 -----YQLVYTKQKEEESQSSSIYHONDITMTPIVTRADPFINNETIAGKDLVAN 646

QY 659 VTAGEFLHPHADIDNTVTVNGVGFELRPYNFEDEPSFYGSADSIYRGDADGACEVN 718
DB 647 VTASFLLHPHADIDNTVTVLGNRVGFLRPYNFEDEPSFYGSADSIYRGDADGACEVN 706
QY 719 PLACLPQAAACAPDLPATFSHGCF 741
DB 707 PVACLPDLAACVDPDLPFSYHGCF 729
RESULT 3
ID ABP41516 standard; Protein; 248 AA.
XX ABP41516;
AC 22-AUG-2002 (first entry)
DT Human ovarian antigen HVBK73, SEQ ID NO:2648.
XX
DE Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; chromosome 7q34-ter.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PT 07-JUN-2000; 2000US-209467P.
XX
PR (HUMA-) HUMAN GENOME SCI INC.
XX
PA Blise CE, Rosen CA;
XX
PI N-PSDB; ABQ54593.
XX
DR WPI: 2002-147878/19.
XX
DR N-PSDB; ABQ54593.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
PS
PS Claim 11; SEQ ID NO 2648; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders


```

XX  Herbicidally active polypeptide SEQ ID NO 1715.
DE
XX  Herbicidal; plant; agriculture; herbicide.
XX
XX  Arabidopsis thaliana.
XX
XX  WO200210210-A2.
XX
XX  07-FEB-2002.
XX
XX  28-AUG-2001; 2001WO-EP09892.
XX
XX  28-AUG-2001; 2001WO-EP09892.
XX
XX  (FARB ) BAYER AG.
XX
XX  Tietjen K, Weidler M;
XX
XX  WPI; 2002-269010/31.
XX
XX  Identifying plant target proteins for herbicidally active compounds,
XX  comprising aligning and comparing nucleic acid or amino acid sequences
XX  from plant with nucleic acid or amino acid sequences from non-plant
XX  organisms -
XX
XX  Claim 5; SEQ ID NO 1715; 261pp + Sequence Listing; English.
XX
XX  The invention relates to identifying target proteins
XX  (ABB92082) for herbicidally active compounds, comprising
XX  aligning and comparing nucleic acid or amino acid sequences from plant
XX  with nucleic acid or amino acid sequences from non-plant organisms using
XX  suitable search parameters, where plant sequences having an E-value
XX  greater by a factor of 3 than the E-value of most similar non-plant
XX  sequences are selected. The polypeptides or nucleic acids encoding them
XX  are useful for identifying modulators. The identified modulators are
XX  useful as herbicides.
XX
XX  Sequence 687 AA:
SQ
Query Match      8.1%; Score 324; DB 23; Length 687;
Best Local Similarity 22.1%; Pred. No. 2.5e-23;
Matches 168; Conservative 102; Mismatches 298; Indels 192; Gaps 31;
OY  20 PHCPSPSPSAQPTWHPGQSLFADLSREELTAVMRFLTORLGLDAAQARPSDNCVFS 79
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  39 PHHP-----LDPLTTP-----EIKRVOTILSGH-DPGFGSGS-----TIIHA 74
OY  80 VELQLPPKAAALHLDGSPPARREALAIVFPGQPOPNV---SELVYVGLPHPSY----- 132
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  75 MALDEPKQKRVIMWKGDRLPRRAETLANSNESHVLTDLKSGRAVSDLVNPTFEGYPI 134
OY  133 --NRDVTVERHGRLPYHRHPRVLEOXYLIDIDOMIFNRELPOASGLHHCFFYKRGNNLV 190
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  135 LTKMDIIV---SQVPI-----KSEFNRSLIARGIF-FSGLICITTFPAGYGPD-- 180
OY  191 TMTTAPRGLOSGDRAVTFGLYINISGAGFLAHV-GLLELVNKAALDPAAMTIOKVEYOG 249
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  181 -----EGRRRVIRIKQFCSKODTVNFMPIEGLYLVDMDKTE----- 218
OY  250 RYDLSLAQLAEQFAGLVNVLVLPDNGTGSWSLSKSPVPGAPAPLPD----- 297
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  219 -----TIKIVDNG-----PVVPPKSTGTETRYVYGLNETVYMDRV 252
OY  298 -----YPOGPRFSVQ-GSRVASSLMTFSFGLAFSGPRIFEDVAFQGERLYVEISLQDAL 350
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  253 NPKMEKPDGSPSFOYEGYLVKAMKFKHIKPDORAGMIT-----SQATVRDCKTGEAR 306
OY  351 AI-YGNSPAA-----TTRYVDG-FGMGKYYTPPLRGVDCPYLATYVDMHFL 397
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  307 SVMKKGFASELFVNMDPGEGWYSKAYMDAGEFGLSPSMPLVPLNDCPRNAYYIDGFFA 366
OY  398 LESQAPRTINDAPCFVFQDNGCLPLRRHSHDLXHYFGGL-----AETVLVYVRMSSTLLN 451

```

```

DB  367 SPEGIPIDLPNMICFERFAGDTSMRHSSEL-----LPQVDINESRAKATLVARMACSVGN 422
OY  452 YDYVMDTVFHPGSAIEIRFYATG-----YISSAFLEFGATGKYGNQVSEHTGLTVH 503
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  423 YDIYFMEFQMDGVIRVTVAAASGMLWKGTAIVENEDLEEKEDDSGPLISENVIQVYH 482
OY  504 SAHFYVDLVAGLENNVMAEDVFPMAVPWSEHLOK---LQYRKLEMEQQAFLV 560
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  483 FISFHLMDIDGSAN---NSFVYHLEKORLPGESEKRSYKVKYAKTEKDQIKM 538
OY  561 GSATPRYLYLAS-NHSNKGHPGRIQMLSFAGEPLPONSSMARGFSMERQOLAVTQK 619
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  539 SLYDPRFHLVNPNNRLSRGNPAGYKLVPGNAAASLDHDDPPQMGCAFTNNGIWT-- 596
OY  620 EEEPPSSSVFNQNDPMA-----PTVD-FSDFINNETIAGDLVAVTAGFLHIP 667
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  597 -----YNRSEQWAGGLMTQSRGEDTLQVWSD--RDRSIEKNDIYLVYTLGFHHPV 645
OY  668 HADIDPNTVTVGNGVGFLLRPYFEDDP-----SFYSAD 702
DB  646 QEDFPVMPITAS--SFELKPVNFESNPVLGISPFEXD 683
RESULT 6
ABB92082
ID  ABB92082 standard; Protein; 759 AA.
XX
AC  ABB92082;
XX
DT  31-MAY-2002 (first entry)
XX
DE  Herbicidally active polypeptide SEQ ID NO 1293.
XX
XX  Herbicidal; plant; agriculture; herbicide.
XX
XX  Arabidopsis thaliana.
XX
XX  WO200210210-A2.
XX
XX  07-FEB-2002.
XX
XX  28-AUG-2001; 2001WO-EP09892.
XX
XX  28-AUG-2001; 2001WO-EP09892.
XX
XX  (FARB ) BAYER AG.
XX
XX  Tietjen K, Weidler M;
XX
XX  WPI; 2002-269010/31.
XX
XX  Identifying plant target proteins for herbicidally active compounds,
XX  comprising aligning and comparing nucleic acid or amino acid sequences
XX  from plant with nucleic acid or amino acid sequences from non-plant
XX  organisms -
XX
XX  Claim 5; SEQ ID NO 1293; 261pp + Sequence Listing; English.
XX
XX  The invention relates to identifying target proteins
XX  (ABB92082) for herbicidally active compounds, comprising
XX  aligning and comparing nucleic acid or amino acid sequences from plant
XX  with nucleic acid or amino acid sequences from non-plant organisms using
XX  suitable search parameters, where plant sequences having an E-value
XX  greater by a factor of 3 than the E-value of most similar non-plant
XX  sequences are selected. The polypeptides or nucleic acids encoding them
XX  are useful for identifying modulators. The identified modulators are
XX  useful as herbicides.
XX
XX  Sequence 759 AA:
SQ
Query Match      7.5%; Score 302.5; DB 23; Length 759;
Best Local Similarity 22.5%; Pred. No. 4.6e-21;

```


XX 05-SEP-1994; 94UP-0211526.
 XX 05-SEP-1994; 94UP-0211526.
 XX (TOKU) TOKUYAMA SODA KK.
 XX WPI: 1996-203152/21.
 DR N-PSDB; AAT14174.
 XX
 PT Gene encoding amine oxidase - used in formation of hydrogen peroxide
 PT from butyl amine, or benzylamine by oxidative deamination
 PS Claim 3; Pages 9-11; 12pp; Japanese.
 CC The present sequence is the A. niger strain M-62 amine oxidase
 CC (AO), which is used in the formation of hydrogen peroxide from
 CC butyl amine, or benzylamine by oxidative deamination. The AO gene,
 CC the sequence of which was determined by sequencing a restriction
 CC enzyme map of an AO gene contg. cDNA, can be used for the
 CC recombinant prodn. of AO.
 XX
 SO Sequence 670 AA:
 Query Match 5.3%; Score 214.5; DB 17; Length 670;
 Best Local Similarity 22.3%; Pred. No. 3,7e-12;
 Matches 168; Conservative 86; Mismatches 301; Indels 197; Gaps 37;
 34 HPQSOLFADLSREBLAVNRFLQRLGPGVLVDAQAARPSDNCVSEVLQLPKA----- 88
 4 HP-----LALISEETIANVILAQHPNTVYIDRE-----IYLSEPPKAQLLEF 48
 89 AALAH---IDRGSPPPARELA---IVFGRQP-----QPNVSELVVGPRLPHS 131
 49 LALEHSGRLSPTSPRPRLALCOVDYIGNDRIPEEESVVDVGRQVQHRVVGKEHHAS 108
 132 YMR---DYVYER-HGGGLPRHRRVYLFQETL-DIDOMIFNRELPOA-----SGL--- 175
 109 LTLSEEDTLVERCFASP-----LFQKALAFED-----LPEGEVVIIEPYGGADYV 153
 176 -----LHHCFFYKHKGRN-----LVYMTTAPRGSGDRATWFGLYYINS 216
 154 EKKRRYFGICFATDKKKNPDANFYSLPLIPVMDGATPGDHSGPSR-----HCKG 208
 217 AGFELHNVGLLELVNHRKALDPARWTIQKVEYQGRYDLSLAQLEAFAGLVNVVLPDNG 276
 209 EGLTEQTFKRDIIGHCDS-----YVELL 234
 277 TGSWSLSKSPVPPGPAPLPLOFYPOGPRFSV-QGSRVASSLWTSFGL---GAFSGPRIF 331
 235 PGGTRKDLKPLN-----VVOPEGSPFRITEESLWEMOKWRFVAENPRRCYHSOTSW 287
 332 DVAFPOGRLVYEISLOEALAIYGSNSPAAMTTRVVD-GFGMGKTYTTPRLRGVCCPLAT 390
 288 ---YDGSVILRLSVSEMTVPYADPRPFHKKQAFDFGSGGGMANNLSIGCCCLLVYIK 344
 391 YVDMHFLLESQAARKTIDATCFVEQONGLPLRRHSDLYSHYEGLAETV---LVYRSM 446
 345 YFAVVMGADGSAKKMNAICLHEQDNGIGMK-----HSMWRGGRVAVTFHRELVAQFI 398
 447 STLNLTYVMDYVFHPSGAIEIRYATGYISSAFLGATGKY---GNQVSE---HTLG 498
 399 ITLANYYEIRAYKFDQSGGIGTGRAVACHGYLE-----RCQHRCWOCQRYVQRQRMRVYG 451
 499 ---TVHHSNHFKVDLVAGLENNVMAEDMVFPMVAPWSEHQLQLOVTRKLEHEEQ 555
 452 PEPPAHLCAHRP---GIYGNNSVQVEESHVPMNAVNTNNGVFYVNT-----ETWER 503
 556 AAFVLSGATP---RXYLIASNH-----SNKWGHPRGYRIQWLSFAGEPLPONSMAFGFS 607
 504 AGFF--DAAPELNLTVMKVNPHKKNPISQK---PVGYKFIPLAQORLLADNLSQARRAQ 558
 608 WERYQLAVTORKEEPESSSVFN-QNDPMAFTVDSDFI-NNETIAGKDLVAVMTAGFLH 665

DB 559 FAQHWWVTKYKRDGEIYAGGRYTLQSQEEIEGV--SDAVRGRGSVDPTDVVWSTFGITH 616
 QY 666 IPhAEDIPNTVTGNGVGFELRPVNFDEPDS 697
 DB 617 NPREVDMF--VMPEVIRQLMIRADFTTAPS 646
 RESULT 9
 AAB42952
 ID AAB42952 standard; Protein; 863 AA.
 AC AAB42952;
 XX 08-FEB-2001 (first entry)
 DT Human ORFX ORF2716 polypeptide sequence SEQ ID NO:5432.
 DE
 XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 KW
 XX Homo sapiens.
 OS
 XX NO200058473-A2.
 PN
 XX 05-OCT-2000.
 PD
 XX 31-MAR-2000; 2000MO-US08621.
 PE
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 XX Shimkets RA, Leach M;
 PI WPI: 2000-602362/57.
 DR N-PSDB; AAC77161.
 DR
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PT
 XX Claim 11; Page 4615-4617; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertoid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

PT Rat choline kinase having specified DNA sequence - useful for
PT diagnosing cranial nerve diseases, liver diseases, tumours, etc.
XX
XX
PS Claim 1, Page 21, 26pp: Japanese.

CC The choline kinase (CK) gene may be obtd. from rat brain, human
CC brain, human liver, chicken liver, egg, soy bean, etc. The rat CK
CC gene is shown, and can be isolated by PCR using oligonucleotides
CC corresp. to partial CK sequences as primers. The oligonucleotides
CC or the amplified DNA fragments, can then be used as probes to detect
CC CK coding sequences. Rat CK may be purified by HPA-agarose affinity
CC chromatography. A large amt. of CK may be produced and antibodies
CC against CK may be obtd. for use in detecting the distribution of CK
CC in tissues or cells, and the diagnosis of cranial nerve disease.
CC liver disease and tumours.
XX
XX

SO Sequence 435 AA;

Query Match 2.7%; Score 109; DB 14; Length 435;
Best Local Similarity 21.5%; Pred. No. 0.11;

Matches 79; Conservative 41; Mismatches 150; Indels 98; Gaps 14;

QY 249 GRYDLSLAQLAOFENAGLVNVLIPDNGTGGSMKSVPECPAPLQFPYQGPFFSVQG 308
D 33 GQORDAAGELSKDGLGSRGSLALPPPPP-----PLPLPPPPSPPLADEQPAPR----- 82
QY 309 SRVASSLTWTFSGFAGFSGPR-----IFDVRFQGERLYVEISLOEALAYGNSPAMTT 363
D 83 TRRAYVLMCKEFLPGAMGLAEDOFHIVINGLNNMLFQCSLPISIVS-GDEPRKYL 141
QY 364 RYVDGFGMGKTYTPPLRGVDCPYLATYVDWHFLESQAPKTIRDAFCVEQ---NQGLP 420
D 142 RL-----YGAIIKMGAEWVLESVMFALIAERSLGP---LYGIFFQGRLEQRIIP 188
QY 421 LRRHSD--LSHYFGGAEVLYVRSMTLLNDYVDYVDFHSGAIEIRFYATGYLSS 478
D 189 SRRDTEELCLPDISAEIAEKMAFTHGKMPFNKEPKM----- 226
QY 479 AFLFGATGKYGNQVSEHTLGTVHTSAHFKYDLDVAGLENNVMADWVFVPAVWSPDH 538
D 227 --LFGTMEKYLNOYLK-----LKFSEARVQQLHK-----FLSYNPLELEN 266
QY 539 QLOQLQVTRK-----LLE---MEQAAFLVGSATPRYLTLA---SNHSN 576
D 267 LRSLLQYTRSFVVECHNDCEGNILLLEGQENSEKOKLMLIDFESSYVNRGFDIGNHFC 326
QY 577 KMGHPRGY 584
D 327 EMTYDTYI 334

RESULT 13
ABG3153
ID ABG3153 standard; Protein: 446 AA.

XX ABG3153;

XX 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #13144.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX
XX (HSE-) HXSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.
XX N-PSDB: AAS77340.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX

PS Claim 20; SEQ ID No 43512; 103pp: English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX

SO Sequence 446 AA;

Query Match 2.7%; Score 106.5; DB 22; Length 446;
Best Local Similarity 21.3%; Pred. No. 0.2;

Matches 65; Conservative 39; Mismatches 98; Indels 103; Gaps 16;

QY 333 VRFGGERLYVEI-SLOELATAYGNSPAMTRY-----YGGGGMGKTYTPPLRGV 383
D 211 LKAEGERAVYQIMQDLKVNMEGDD---LTRYGQKYNKLTLLHGFGFSCPNRDGTIRG- 265
QY 384 DCPY--LATYVDHMFLESQAPKTIRDAFCVEQDNQGLPLRRHSDL-----YSHFGG 435
D 266 GCTFCNVASPAD-----EAQHRSAIEOLA-----HOANLVNRAKRYLGTFQA 308
QY 436 LAET--VLVVRSMSTLLNDYVDYVDFHSGAIEIR-FYATGYISSAF-----LFGAT 485
D 309 YTSFAEYQVLRSM-----YQAAVEHTGKVGITGFCGGSVNAAYAVPELMCAV 360
QY 486 GKYNQVSEHTLGTVHTSAHFKYDLDVAGLENNVMADWVFVMAVWSPDHQLOQLQV 545
D 361 PFYGRQAPFADVAKTEAPLLHFAELDRINEG-----PAYE-AALKA 403
QY 546 TRKLEMEEQAAFLVGSATPRYLTLASNSHMKWCHPGRYRIOMLSFAGEPLPOJSSMARG 605
D 404 NNKYYE-----AYITPGYNH-----GFHNDSTPRIDKSAAD 434
QY 606 FSWER 610
D 435 LAMOR 439

RESULT 14
AAE23407
ID AAE23407 standard; Protein: 617 AA.

AC AAE23407;
 XX
 XX 27-AUG-2002 (first entry)
 DE Lysine oxidase proenzyme mut3.
 XX
 XX Amino acid oxidase; insect infestation; gene therapy; boll weevil; BWV;
 KW corn rootworm; CRM; insecticide; wireworm; WM; Colorado potato beetle;
 KW CPB; lysine oxidase; tetanactam synthase; enzyme; E.C. 1.4.3.2;
 XX mutcin; mutant.
 OS Unidentified.
 XX
 XX Key Location/Qualifiers
 FH Region 68..86
 FT /note= "Protease sensitive region"
 FT
 FT Misc-difference 70
 FT /note= "Wild-type Ala substituted with Gly"
 FT 71
 FT Misc-difference 71
 FT /note= "Wild-type Leu substituted with Gly"
 FT 72
 FT Misc-difference 72
 FT /note= "Wild-type Leu substituted with Gly"
 FT 73
 FT Misc-difference 73
 FT /note= "Wild-type Lys substituted with Gly"
 FT 74
 FT Misc-difference 74
 FT /label= Unknown
 FT /note= "Wild-type Glu substituted with Xaa"
 FT 75
 FT Misc-difference 75
 FT /label= Unknown
 FT /note= "Wild-type Ala substituted with Xaa"
 FT 76
 FT Misc-difference 76
 FT /label= Unknown
 FT /note= "Wild-type Pro substituted with Xaa"
 FT 77
 FT Misc-difference 77
 FT /label= Unknown
 FT /note= "Wild-type Arg substituted with Xaa"
 FT 78
 FT Misc-difference 78
 FT /label= Unknown
 FT /note= "Wild-type Ala substituted with Xaa"
 FT 79
 FT Misc-difference 79
 FT /label= Unknown
 FT /note= "Wild-type Glu substituted with Xaa"
 FT 80
 FT Misc-difference 80
 FT /note= "Wild-type Glu substituted with Gly"
 FT 81
 FT Misc-difference 81
 FT /note= "Wild-type Glu substituted with Gly"
 FT 82
 FT Misc-difference 82
 FT /note= "Wild-type Leu substituted with Gly"
 FT
 XX US6372211-B1.
 PN
 XX 16-APR-2002.
 PD
 XX 21-APR-1998; 98US-0063733.
 PF
 XX 21-APR-1997; 97US-044504P.
 PR
 XX (MONS) MONSANTO TECHNOLOGY LLC.
 PA
 XX Isaac BG, Greenplate JT, Purcell JP, Romano CP;
 PI
 XX WPI; 2002-424731/45.
 DR
 XX
 XX Composition for controlling coleopteran insect infestation of plants,
 PT such as Diabrotica, Melanotus, Leptinotarsa, or Anthonomus infestation,
 PT comprises a lysine oxidase enzyme and a tetanactam synthase enzyme -
 XX
 XX Example 13; Column 89-94; 77pp; English.
 XX
 XX The invention relates to composition and methods for controlling
 CC coleopteran insect infestation of plants, by co-expressing an amino
 CC acid oxidase and a second enzyme that provides insecticidal activity
 CC when present in a mixture with the amino acid oxidase. Nucleic acid

CC sequences encoding these enzymes are used in gene therapy. The
 CC composition is used to control insect infestation of plants. It is
 CC used to control coleopteran species Diabrotica, Anthonomus, Melanotus,
 CC or Leptinotarsa. It is especially used to control boll weevil (BWV),
 CC corn rootworm (CRM), wireworm (WM) or Colorado potato beetle (CPB).
 CC The composition can also be used to transform plants that can express
 CC the enzymes of the composition. The present sequence is a lysine
 CC oxidase (E.C. 1.4.3.2) proenzyme mutant used in the exemplification
 CC of the invention.
 CC
 XX
 XX Sequence 617 AA;
 SQ
 Query Match 2.6%; Score 106; DB 23; Length 617;
 Best Local Similarity 22.8%; Pred. No. 0.38;
 Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;
 44 LSREELTAVMRFTLQRLG--PGLVDAQAAR-----PSDNCVFSVELQLPPKAA 90
 18 LIREKAKELNITLRLGEPGIPPNESGFLGGYSHDLPSDPLYSISK----PGGG 73
 91 LAHLDRGSPPPAREALAIYFFGRQPPNSELVVG-----PLPHSYMDVVERHGR 144
 74 XXXXXXGGGPPRK--VCIYAG-----VSGLYIAMLLDLKIPNLYDIFESSSRTRG 125
 145 LPYHRRPVLFGQYLDIDDMFNRRLPOAGSLHHCFCYKRGMLVMTTAPRGLSGDR 204
 126 LYTHHTDAAKHDDYDIGAMV-----PDPSMKRTFNLFKRTGAPLI-----KYLDGEN 175
 205 ATWFGLYYNISGAGFLLHVGLELVN-----HKALDPARMTIQKVFYGRYDLSLA 256
 176 TP--QLXNN---HFFAKGVSDPYMVSANGTVDDVDVSGEKLQOAR--GYKEKLA 227
 257 QLEAGFAGLVNVLIPDNLTGSGWSLKSVPYPPGAPPLDIFYOGPFVSQGSRYVASSLW 316
 228 E--DFDKGFDELALVDMDMT--REYLKRGPGKEARKYDF-----FAIQMMETON-- 273
 317 TFSFGUGAFSGPRIFDVRFQGERLVEISLQ-----PALAIYGNS 357
 274 -----TGTNLFDAQF--SESVYDSFDDNPTRKPMWYCLEGSTS 309
 Db
 RESULT 15
 ABB67526
 ID ABB67526 standard; Protein; 825 AA.
 XX
 AC ABB67526;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 29370.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN W0200171042-A2.
 PD
 XX 27-SEP-2001.
 PF
 XX 23-MAR-2001; 2001WO-US09231.
 PR
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 PI
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL11629.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 12:02:34 ; Search time 20 Seconds

(Without alignments)
1094.533 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763

Perfect score: 4010
Sequence: 1 LVCVLLVGRGSGDEPSQLP.....OMACAPDLPASHGFSHN 744

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	2.6	617	4	US-09-063-733A-58
2	102.5	2.6	541	4	US-09-158-767-19
3	102.5	2.6	541	4	US-09-158-767-20
4	101	2.5	617	4	US-09-063-733A-46
5	100.5	2.5	1253	4	US-08-506-296B-14
6	100	2.5	617	4	US-09-314-242-2
7	95	2.4	617	4	US-09-063-733A-57
8	94.5	2.4	1751	4	US-09-135-574A-44
9	94.5	2.4	15281	2	US-08-471-119A-2
10	93.5	2.3	485	5	PCT-US94-03437-8
11	93.5	2.3	486	1	US-07-737-071A-2
12	93.5	2.3	486	1	US-08-022-096-2
13	93.5	2.3	486	3	US-08-445-463B-8
14	93.5	2.3	486	3	US-08-445-463B-8
15	93.5	2.3	486	3	US-08-445-463B-8
16	93	2.3	2972	4	US-08-044-857D-8
17	93	2.3	3118	4	US-09-579-181-1
18	92.5	2.3	448	2	US-08-078-311-12
19	92.5	2.3	448	2	US-08-460-402-12
20	92.5	2.3	493	2	US-08-078-311-7
21	92.5	2.3	493	2	US-08-460-402-7
22	92.5	2.3	635	1	US-08-184-327A-4
23	92.5	2.3	635	2	US-08-078-311-1
24	92.5	2.3	635	2	US-08-460-402-1
25	92.5	2.3	635	5	PCT-US95-00670-4
26	90.5	2.3	1436	2	US-08-652-971-2
27	90.5	2.3	1436	2	US-08-991-258A-2

28	90.5	2.3	1436	2	US-08-769-399-2	Sequence 2, Appl
29	90.5	2.3	1436	3	US-08-991-953A-2	Sequence 2, Appl
30	90.5	2.3	1467	4	US-09-523-656-38	Sequence 38, Appl
31	90	2.2	617	4	US-09-063-733A-56	Sequence 56, Appl
32	89.5	2.2	623	1	US-08-332-838-2	Sequence 2, Appl
33	89.5	2.2	630	4	US-09-228-986-71	Sequence 71, Appl
34	89.5	2.2	911	4	US-08-460-269C-4	Sequence 4, Appl
35	89	2.2	688	4	US-08-973-273-26	Sequence 26, Appl
36	89	2.2	747	4	US-08-973-273-3	Sequence 3, Appl
37	88.5	2.2	379	3	US-09-028-934-36	Sequence 36, Appl
38	88.5	2.2	512	1	US-08-462-484-6	Sequence 6, Appl
39	88.5	2.2	512	1	US-08-441-147-6	Sequence 6, Appl
40	88.5	2.2	512	5	PCT-US95-07536-6	Sequence 6, Appl
41	88.5	2.2	583	4	US-09-401-476-2	Sequence 2, Appl
42	87.5	2.2	353	4	US-08-838-151A-49	Sequence 49, Appl
43	87.5	2.2	485	3	US-08-445-463B-2	Sequence 2, Appl
44	87.5	2.2	485	3	US-08-445-463B-2	Sequence 2, Appl
45	87.5	2.2	485	4	US-08-044-857D-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-09-063-733A-58
; Sequence 58, Application US/09063733A
; Patent No. 6372211
;
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Arnold White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,733A
; FILING DATE: 21-APR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-787-1440
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
;
; US-09-063-733A-58
;
; Query Match 2.6%; Score 106; DB 4; Length 617;
; Best Local Similarity 22.8%; Pred. No. 0.071;
; Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;
;
; 44 LSEELTAMRFLTOLG--PGVDAQAAR-----PSDNCVFSEVLQPKAA 90
; 18 LIREKAKELNITLERGEVPIPPNEGFIAGGYSHDNLPSDPLYSTIK---PGGGG 73
```

QY	91	LAHLDRGSPPAKALALIVFFGRRPOVNSVELVVG-----PLPHSPYMDVYVERHGCF	144
Db	74	XXXXXXXXGGGGRPK--VCIVGAG-----VGLYIAMIIDLKLPINLYTDFESSSRFGGR	125
QY	145	LPYHRRPVLFQGEYLDIDIOIMIFNRELPOASGLHHCFFYKRGHNLVYMTTAPRGLSGDR	204
Db	126	LYTHHFPDAKHNDYDIDIGAMRY---PDIPSMKRFNFLEKRTGMPLI-----KYLDGEN	175
QY	205	ATWFGGLYNINISGAGFFLIHHGLELVN-----NKALDPARKTIOKVYQGRYIDSIA	256
Db	176	TP--QLYNN---HFFAKGSVDPRYMSVANGTVPDDVDSVSGEKLQDAF--GYKKKLA	227
QY	257	QLEAQFEAGLVNVVLLIDNGTGGGSMSEKSPVPGPAPLPQFYPOGPRFSYVSGSRVASSLM	316
Db	228	E--DFDQGFPELMLVYDMMT--KEYLKRGRGPKAERKYDE-----FALQMMETQN---	273
QY	317	TFPSGLGAFSGPRIFDVRFOGERLIVETISIQ-----EALAIYGGNS	357
Db	274	-----TGTNLFQDAF--SESVYIDSFDPDNPRTKPMWYIEGGTS	309

```

RESULT 2
US-09-158-767-19
: Sequence 19, Application US/09158767A
: Patent No. 6180363
: GENERAL INFORMATION:
: APPLICANT: Batard, Yannick
: APPLICANT: Durst, Francis
: APPLICANT: Schalk, Michel
: APPLICANT: Werck-Reichardt, Daniele
: TITLE OF INVENTION: RECODING OF DNA SEQUENCES PERMITTING
: TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST
: FILE REFERENCE: A32000
: CURRENT APPLICATION NUMBER: US/09/158,767A
: CURRENT FILING DATE: 1998-09-23
: EARLIER APPLICATION NUMBER: FR 97-12094
: EARLIER FILING DATE: 1997-09-24
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 19
: LENGTH: 541
: TYPE: PRS
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: a)tered sequences
US-09-158-767-19

```

Query Match 2.6%; Score 102.5; DB 4; Length 541;
Best Local Similarity 19.6%; Pred. No. 0.13;
Matches 109; Conservative 60; Mismatches 205; Indels 183; Gaps 22.

```

0Y 151 PVLPEVYDIDOMIFENRRLPOASLLHNCCP-----YHNCRNLYUTWMTAPRGJOSGDBRAT 206
Db 40 PELVHADMEHMEWJAG-NLRNAGGTOTCFIAPNPGVARRG-GLVUTYCDPRNLEHVLKAR 97
0Y 207 -----WGLYVYNISGAGFF-----LHNHGLELLVNHKALDEARWTKOVFY 247
Db 98 FDNYPKPGFPMHGVFEDLLDGDIFNSDGDWTLAOKRTALEFTTETRLTAMSRYVRSRI-- 155
0Y 248 OGRYVDSLALQ---EAQENAGLVVWVLIIPDNGTGSMSLASPVPRGPAPPLQYPOGPRF 304
Db 156 HGRLLPILADAKKAKRAOVDLDDLRLRLFDNIGCLAG---KDPETLAQGLPEHEFA 209
0Y 305 SVQSGRVASLSMTFSF-----GLGAFSGPRIEDVRFQGERLVEYSIQEALAIY 353
Db 210 SAFEDRATFATLNRFIFEPFLRCKKWLGLG-----METILTSSMAHY 251
0Y 354 GGNSPAAMTTRVYDVGPFQMGKTYTPLR-----CYDC 385
Db 252 DQYLAIVATIKRRLELAAGKCDPATHTDILLFSFMKRGYSDESLQHVALLNPLAGRDT 311
0Y 386 PYLATYVWHEHLESQAP-----KTIRDAFCVFEQNOGL-----PLRRHSDILTSY 432

```

Db	312	SSVA- -LSMFVLVSTHRAVERKIVRELCSVLASRBAHPALMLAPFEEELDRVYL	369
Qy	433	EGGAEIYLVVRSM- -STLNDIYMDVYFHHSG- AIEIRPATGYISSALPLGATG	487
Db	370	KAAISETIRLPSVPEDSKHVADDYLPDGTFTVAGSSVTSYISAGRMG- -	420
Qy	488	YGNVSEHTLCTVTHSHAFKVDYDVLGLENMVAEDMVEPMAVPWSPEHOLQVTR	547
Db	421	- - - - -VNGED- - - - -CLEFRPE- - - - -RMLSADG	439
Qy	548	KLEMEQOAFVLGSAAPRYLYLASNSHMKWGHPRGYRIOLSPFAGPELPONSSMARGFS	607
Db	440	TKFEHDSDIKYFAVFNAGRYVL- - - - -GRDLAY-LQMKNTAGSVLLRH- - - - -	481
Qy	608	WERIOLAVTORKEEPS	624
Db	482	- - - - -RLTVAGHREOKMS	496

```

RESULT 3
US-09-158-767-20
; Sequence 20, Application US/09158767A
; Patent No. 6180363
; GENERAL INFORMATION:
; APPLICANT: Batard, Yannick
; APPLICANT: Durst, Francis
; APPLICANT: Schalk, Michel
; APPLICANT: Werck-Reichardt, Daniele
; TITLE OF INVENTION: RECORDING OF DNA SEQUENCES PERMITTING
; TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEASTS
; FILE REFERENCE: A32000
; CURRENT APPLICATION NUMBER: US/09/158,767A
; CURRENT FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: FR 97-12094
; EARLIER FILING DATE: 1997-09-24
; NUMBER OF SEQ. ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 541
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Altered sequences
US-09-158-767-20

```

Query Match	2.6%;	Score 102.5;	DB 4;	Length 541;
Best Local Similarity	19.68;	Pred. No. 0.13;		
Matches 109; Conservative	60;	Mismatches 205;	Indels 183;	Gaps 22.

```

Qy 151 PPFPEYLDIOMJFNRLRPASGLHNHC-----YHNGBNLTMTTAPRGLSGDSBRAT 206
Db 40 PELVQHAEDMEHMLG-LNRRAGGTIOTCIATVAGVARG-GLVTVICDPRNLEHULKAR 97
Qy 207 -----WGLYNYNISAGF-----LHHVGLLELVNKKALDPAKMTIOKVEY 247
Db 98 FPNVKEGFGHGVFEDLLDGDJFNDSGDVTLAOKRTALETTTPTTLTAMSRYWSRSI-- 155S
Qy 248 OGRYVDSLALQL---EAPREAGLVNVVLLPNDGTGGSMSLKSPVPRGAPRLQGFYQSPFR 304S
Db 156 HGRLLPILADAKKAGQVDDLDLLRLFLDNICGLAG-----KDETLAQGLPENEFA 209S
Qy 305 SVQGSRVASSLMTSEF-----GLGAFSGPRIFDVAFQGERLYEISLQALAIY 353S
Db 210 SAFDATEATLRLRFPEFLMCKKWLGLG-----METTLTSSMAHY 251S
Qy 354 GANSPPAAMTRVDDGFGMCKYTPPLR-----GVDC 385S
Db 252 DQYLAIVIKKKRLLELAANGCKDPAATHDDLRSFMRKGSVSDSLQHVALLNTFLAARDT 311S
Qy 386 PLYATIVDMHFLLESQAP---KTRDAFCYFEONQGL-----PLRHHNSDLYSHY 432S
Db 312 SSVA--LISGFVLTSTHAPERKVIYRELCSVLASRGNDPDMALMAPPFEELDRIVYL 369S

```

```

QY 488 YGNVSEHTLGTVHTSHAHFKVDLDVAGLENNWVAEDMFEVPAWVPSPEHOLQLOVTR 547
Db 421 -----WGEO-----CLERPRE-----RWLSADG 439
QY 548 KLEMEQALVGSATPRVLYLASNHSNKWGRGRYIOMLSRAGEPLQONSSMARGFS 607
Db 440 TKFQHOHSYKRVAAENACPRVCL-----GKDLAY-LQMKNTAGSVLLRH----- 481
QY 608 WERYOLAVTORKEEPS 624
Db 482 --RLTVAPGRHVEQKMS 496

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

OY      203 DRATFGLLYNISCAGFFLHHVGLLELVN-----HKALDPARWTTQKVYFOGRYDS 254
          :      :      :      :      :      :      :      :      :      :
Db      174 ENTP--OLLYN-----HEFAKGVSPRYVAVANGTVEDDVDVDSYGEKIOOAF--GYKKEK 225
          :      :      :      :      :      :      :      :      :      :
OY      255 LAOLEAQEAGLVNVLIPDNGTSGMSLKSPPVGPAPPLQFQPOGPRPSVSGSRVASS 314
          ||      :      :      :      :      :      :      :      :      :
Db      226 LAE---DDDKGEDELMLVDMTT--REYLKRGKGPKEAPRYDF-----FALOMETON- 273
          ||      :      :      :      :      :      :      :      :      :
OY      315 LMTFSGIGAFSSGPRIDVRFQOGERLYEISLQ-----EALAIYGSNS 357
          :      :      :      :      :      :      :      :      :      :
Db      274 -----TGTLNLFDOAF--SSSVYIDSPFDNPPTKPEWVCIGGNS 309

```

```

RESULT 5
US-08-506-296B-14
; Sequence 14, Application US/08506296B
; Patent No. 6313265
GENERAL INFORMATION:
APPLICANT: Phillips, Greg
APPLICANT: Cunningham, Bruce A.
INVENTOR: Crossin, Kathryn L.
TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute
STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
CITY: La Jolla
STATE: California
COUNTRY: U.S.
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506/296B
FILING DATE: 24-JUL-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-506-296B-14
Query Match 2.5%; Score 100.5; DB 4; Length 1253;
Best Local Similarity 18.5%; Pred. No. 0.91;
Matches 160; Conservative 89; Mismatches 269; Indels 349; Gaps 40;
QY 4 VLVVRGGDG-----GEPSLPHCPVSFSACP-WTHPQSOLFA----- 42
: || | | : | : | : | : | :
DB 299 ILKVEEDDDRCIAENSLGSARHAYVTVYAADYLHKPKDSHLVGGETARDLCOVQG 358
: | | | | : | : | : | : | :
QY 43 -----DLSEKETAVNRFLTORLGGLVDNAOARSNDNCVFSE----- 81
: | | | : | : | : | : | :
DB 359 RQPETWIRINKIPPELAELAKOKRYIQR---CALLLSNVGPSDFIVVTCEARNRHGLLA 415
: | | | : | : | : | : | :
QY 82 -----LDLPKKAAALAHIDRGSPPARDEALATVEFGKPQOPNMSVLVGPLPPIHWKD 135
: | | | : | : | : | : | :

```



```

Db 416 NAVIYVQIPAKILITADNOTYMAVAGSTAYLLCKAFG-APVPSV----- 458
QY 136 VTVRHGGPPLPHRRPVLFOEYLDIDOMIF---NRELPOASGLLHCCFYKRGKRNLYTM 192
Db 459 -----QWLEDEDTTVLQDERFEFFYANG----- 480
QY 193 TTAARGIQRGATWFGIYUYNISGAGFELHHVLELLVNHALDPAWITQKVYQGRY 252
Db 481 TLGRIDLOANDTGVYFCLANDON-----NTIMANKVKATQIT----- 521
QY 253 DSLAQLAEAGELVNVLLIPDNGTGSWSLKSFPQAPPLQFYPOGPRFSVQ--GSR 310
Db 522 -----QGRSTIEKKGR 534
QY 311 VASSLMTFSGIAGFSGRIFD-----VRFQGE-RLVYEISLOBALAIYGNSPAAMTT 363
Db 535 V-----TFQOASDPSPLOPSTWRGDRDGLGSDKXYFIEGGR-VIHSI 581
QY 364 RYVDGFGMGKYTPPLTRGVDCPYLATYVDHMFLESQA-----PKTRIDAFV 412
Db 582 DYSD-----QGVYSCVASTELD-----VVESRAQLLVGSPQPRRLVLSDLH 625
QY 413 FEQVQ-----GLPLRHHSDLYSHFGIETVLYVRSKSTILNDYVDVIFHPSGALEI 468
Db 626 LTOQOVHVSMSPADHNAPIEK-----YDIEFEDEKEMAPE 660
QY 469 REYATG-----YISSAFLEGATGKYG---NOVSEHMLGTVHT-HSAHF 508
Db 661 KMYSLGKVPNGQSTTLKISPYHYTRVYAIINKYGQEPSPVSE---TYVTEPAAREK 716
QY 509 VDLDAVAGLEMMVADNVFVPMVPMSP-----EHQLQ-RLQVTRKLEMEQA 556
Db 717 NPVDVKG-----EGNETTNMVTITMKPLRMMDNAPQVQRYVQMRPGQTG--PMQOI 767
QY 557 A---ELVGSATPRRLYLASHNSNKGHPRGRIOMLSFAGE-PLPONSNAKRGSMERYQ 612
Db 768 VSDPFLVASTSTFV-----PYELIKVQVNSQGGPEPO---VYIGYSGEYR 812
QY 613 LAYTORKEEPESSSVFNQNDPAPTYVDSFINNETIAGKDLVAMTAG-----FLH 665
Db 813 QAIPELGEIELNNSAVLVK--WRP-VDLAQVKGH--LRGINVYVWEGSORKSKRHIIH 867
QY 666 IPHAEDIPNTVYVNGVGFRLRPNFDEDPSEFYSADSIYFRGDODA-----G 713
Db 868 KDHYVVPANTTSV---ILSGIARPSYHLEVOAENG---RSGSPASEFTFSTPEGVPG 919
QY 714 ACEVNPILACIPQAAACAPDLPAPSHGC 740
Db 920 HPEALHLECOSNTSLLLRMOPPLSHNG 946

RESULT 6
US-09-314-242-2
; Sequence 2, Application US/09314242A
; Patent No. 6248575
; GENERAL INFORMATION:
; APPLICANT: Elizabeth J. Golightly
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; TITLE OF INVENTION: Having L-Amino Acid Oxidase Activity
; FILE REFERENCE: 5556,200-US
; CURRENT APPLICATION NUMBER: US/09/314,242A
; EARLIER FILING DATE: 1999-05-18
; EARLIER APPLICATION NUMBER: 09/080,428
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Trichoderma harzianum
US-09-314-242-2
Query Match 2.5%; Score 100; DB 4; Length 617;

```

```

Best Local Similarity 23.9%; Pred. No. 0.3; Mismatches 83; Conservative 36; Mismatches 139; Indels 90; Gaps 18;
QY 44 LSRELTAVMRFITQRLG--PGLVDAQAAR-----PSDNCYSEYELQPPRAA 90
Db 18 LIREKAKELNILLERLGEVPGIRPPNDRGLGGGSHDNLPSPILSSI-----KPL 71
QY 91 LAHLDRGP--PPAREALAIYFGRQPOPNVSELVVG-----PLPSPYRQDYVERHG 142
Db 72 LKEAPRAEEELPPRK--VCIVGAG-----VSGIYIMILDLKIPILTYDIFESSRTG 123
QY 143 GPLPYHRRPVLFOEYLDIDOMIENRELPOASGLLHCCFYKRGKRNLYTMATARGIQC 202
Db 124 GRLYTHHTDAKHDIYDGAARY-PDIPSKM-----KFNPLKRTKMPILYLYL 171
QY 203 DRATWFGIYUYNISGAGFELHHVLELLVN-----HKALDPAWITQKVYQGRYDS 254
Db 172 DGENTPOLYNN---HFEKQVSDPYVAVSVANGTVPDDVDVDSVGEKQAQAF--GYKKE 225
QY 255 LAQLEAGFAGLVNVLLIPDNGTGSWSLKSFPVGPAPPLQFYPOGPRFSVQGRVASS 314
Db 226 LAE---DEKGFDEMLVDDMTT--REYLKRGKPGKGAAPKYDF-----FAIQMETON- 273
QY 315 LMTFSGIAGFSGRIFDVRFOGERLYEISLQ-----EALAIYGNS 357
Db 274 -----TGTNLFDQAF-SESVIDSFDFDNPTKPEWTCIGGTS 309

RESULT 7
US-09-063-733A-57
; Sequence 57, Application US/09063733A
; Patent No. 6372211
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,733A
; FILING DATE: 21-Apr-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-787-1440
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
US-09-063-733A-57
Query Match 2.4%; Score 95; DB 4; Length 617;
Best Local Similarity 22.5%; Pred. No. 1;

```

Matches 78: Conservative 40: Mismatches 142: Indels 86: Gaps 18:

QY 44 LSRELLTAVMRELTOGLG--PGLVDAQAAR-----PSDNCVSEVELQLPKAA 90
 18 LIREVAKMLNLTRELGEVPGIPPPNEGRLGGSYHNDLPSDPLYSI-----GGSGG 73
 QY 91 LAHLDRGSPPARREALAIVFEGROPQPNVSELVVG-----PLPHPSYKRDVTVENHGP 144
 Db 74 XXXXXGGGPPRR--VCYIAGAG-----VSGLYIAMILDLKIPNLTVDIFFSSSRRTGR 125
 QY 145 LPHRRPVLFOEYLDIDOMIFNRELPOASGLLHCCFYKHKGRGNLTMTTAPRGLOSGDR 204
 Db 126 LTHHTFDKADHDYDIDGAMRY-----PDIPSKRTFNLFRKTGAPLI-----KYIDGEN 175
 QY 205 ATWFGLYYNISGAGFLHHVLELVN-----HKALDPAWMTIQKVFYOGRYVDSLA 256
 Db 176 TP--QLYNN--HFEAKGVSDPYWYVANGSVDPDDVDVDSVGEKLQDAF--GYREKLA 227
 QY 257 QLEAOFAGLVNVLIPDNGTGSMSLSPVPPGAPPLQFTPOGREFSVQGSRVASLSL 316
 Db 228 E--DEDKGFDELMVDMDTT--REYLKRGKGPKEAPKYDF-----FAIQWMETON--- 273
 QY 317 TFSFGAFSPRIIPDFRFOGERLVEISLQ-----EALAIYGN 357
 Db 274 -----TGTLNLPDAF--SESVIDSFDPDNTRKPEWICIEGTS 309

RESULT 8
 US-09-136-574A-44
 : Sequence 44, Application US/09136574A
 : Patent No. 6294366
 : GENERAL INFORMATION:
 : APPLICANT: Farrington, Graham K.
 : Anderson, Paige
 : Gibbs, Moreland
 : Bergquist, Peter
 : Daniels, Roy
 : Morgan, Hugh W.
 : Williams, Diane P.
 : TITLE OF INVENTION: Compositions and Methods for
 : Treating Cellulose Containing Fabrics Using Truncated
 : Cellulase Enzyme Compositions
 : NUMBER OF SEQUENCES: 49
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Howson and Howson
 : STREET: Spring House Corporate Center, P.O. Box 457
 : CITY: Spring House
 : STATE: PA
 : COUNTRY: USA
 : ZIP: 19477
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/136, 574A
 : FILING DATE: 19-Aug-1998
 : CLASSIFICATION: <Unknown>
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/932,571
 : FILING DATE: September 19, 1997
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Bak, Mary E.
 : REGISTRATION NUMBER: 31,215
 : REFERENCE/DOCKET NUMBER: 1997US001/CIP
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 215-540-9200
 : TELEFAX: 215-540-5818
 : TELEX: <Unknown>
 : INFORMATION FOR SEQ ID NO: 44:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1751 amino acids
 : TYPE: amino acid

STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 44:
 US-09-136-574A-44

Query Match 2.4% Score 94.5; DB 4; Length 1751;
 Best Local Similarity 19.0%; Pred. No. 6.9;
 Matches 139; Conservative 83; Mismatches 284; Indels 225; Gaps 35;

QY 10 GGDGGEPSQLPHCPSPSPAPW---THPGQSOLFADLSRELLTAVMRELTOGLGGLVD 66
 Db 578 GGEVEHKQAQAKRISV--PGYPMDFPNDPSYKGLTQLEKNKYIA----- 621
 QY 67 AAQARPSDNCVSVSELQLPKAAALAHLDGSPPARREALAIVFEGROPQPNVSELVGP 126
 Db 622 ---AYDNMLVWGLE---PGAATSTPAPTSTPTPT-----PTPTPTVATPT-P 663
 QY 127 LPHPS-----YMRDVTYERHGGPLPHRRPVLFOEYLDIDOM-----IFN 166
 Db 664 TPTPTPTGSPGTGSGYKVLKNNETSASTGISRPWKIYNGSGSSVDLERKIRIYTYVD 723
 QY 167 RELPOASGLLHCCFYKHKGRNLVTMTTAPRGLOSGDRATWFGLYYNISGAGFLHHVGL 226
 Db 724 GDNFQSA-----VCDMAQIGASNVTFNFK--LSSG-----VSGADYTL-EVGF 764
 QY 227 ELLVNHKALDPAWMT--IQKVFYOGRY-----YDLSLAOLEAOFAG-----LVNVL 272
 Db 765 SSGAGQ--LQPGKDTDIDQVRFKNDMSNYNQADMSWLOSMNYGENAKVTLYVGVLY 822
 QY 273 ----PDNGTGSMSLSPV-----PGAP-PLOFTYPOGREFSVQGSRVASLSMTFSF 320
 Db 823 WGEPPGAPPTAPTSTPT 882
 QY 321 GLGAFS-----GPRIFVRFQGERLVEI--SLOEALAIYG 354
 Db 883 SYGALKVWYANGMLSSPTVNLNPKIKIENNGTAVDSRKVYVWYTIQDEAQSIVAS 942
 QY 355 GNSPALMTTRYVDGFGMGKTYTPLTRGVDCPYLATIYVDMH-----LLESAPKTI 407
 Db 943 SINPAVIDVFKVLGANAG-----GAD--YVEIGFSGAGVLAAGOSTKEIRL 989
 QY 408 ---DAFCVEQNOGLPRRHSDLSYHFGLAETVLYVMSMTLNTYDVMTPTVHPSG 464
 Db 990 SIQKSGSINQNDYVRSANSTIENKTYGIDV-----YKGR--EPGR 1034
 QY 465 AIEIR-FYATGYISSAFLFGATGKYGNQVSEHTLGTVTHSAHKYDLDAAGLENNYMAE 523
 Db 1035 NAOIKWYANGNIGS-----MTVNLNPKIKIENNGT-----AVDSVKKRYWYTI 1082
 QY 524 DMVFVPMVWSPBHOLQLOVTRKLLMEMEQALVGSATPRYL--YLASHNSNKKRGP 561
 Db 1083 GEA-----TOSVSYTSINPAVIDVKFKLGNANGGAD 1115
 QY 582 RGRIGQLSFRAGEPLPONSSMARGFSWERYQLAVTQKKEEPSSSVFENQNDPAPTVD 641
 Db 1116 YVEIIGKSGAG-----VLAGQSTKEIRLSIQ-----GSGSINQNDYVRS 1160
 QY 642 SDFINNETIAG 652
 Db 1161 TGYIENKXYG 1171

RESULT 9
 US-08-471-119A-2
 : Sequence 2, Application US/08471119A
 : Patent No. 5827706
 : GENERAL INFORMATION:
 : APPLICANT: Lettner, Ernst
 : APPLICANT: Schneider, Elisabeth
 : APPLICANT: Schoergerdorfer, Kurt
 : APPLICANT: Weber, Gerhard
 : TITLE OF INVENTION: Cyclosporin Synthetase

```

? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: NO. 5827706artis Corporation
? STREET: 59 Route 10
? City: East Hanover
? STATE: New Jersey
? COUNTRY: USA
? ZIP: 07936
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/471,119A
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Kassenoff, Melvyn
? REGISTRATION NUMBER: 26,389
? REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201 503 8474
? TELEFAX: 201 503 8807
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 15281 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: unknown
? MOLECULE TYPE: protein
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Tolypocladium niveum
? STRAIN: ATCC 34921
?
US-08-471-119A-2

Query Match      2.4%; Score 94.5; DB 2; Length 15281;
Best Local Similarity 18.1%; Pred. No. 2.9e+02;
Matches 154; Conservative 110; Mismatches 307; Indels 279; Gaps 41;

QY 26 SPSPQPTHTGQSGOLPQADSLREELTAVMRLTORL-GPGLVDAQAAPSDNCVSEVLQ 84
DB 13682 NPDLATVDTSTKLTYALDQSDHVAAMLSKQLPAESIYVVLAPRSSEIVACIGI-- 12739
QY 85 PKAAALAHIDRSPRAREALAIIFGGRQPPRVSELVGLPLPHPSYMRDVTYERHGGP 144
DB 12740 --LKANALATIPMDSNVEARQAAL--SEIPGEKFLVGA-----GVP 12778
QY 145 LPYHRRFVLFOEYLDIDIMIFNRELPOASGLLHHCFCYKHNRLVTMTTAPRGLQSGDR 204
DB 12779 IPDKT-----ADV-RMVFISD-----IYASKTDKSYSGTR 12809
QY 205 ATWFGLYYN--SGA-----GFLHNVGLLEL-----VNHKALDPARMT 241
DB 12810 PSASSLAYVIFTGSGTGRPKGVMEHGRVISTLYKONASRIPOSTLMAHVSNLAFDASWE 12869
QY 242 IQKVFQGG-----RYX-----DSLAEQAEAGLVNVLLIPDNTGGSWSLSKVPFPGP- 291
DB 12870 IFTTLNGGLFCISFTVLDKRA--LSAASFSDHRINITLLPRA-----LLKQCLADAPS 12922
QY 292 -----APPLQFYPOGPRFSYOG-----SRVASSLW-----PFSFG--L 322
DB 12923 VLSSLESLEYIGDRIDGADMTKVKDLVKKAYANAGPTENSVMSTIYTIHEHTFANVPVI 12982
QY 323 GASSGR-----IDVRFQGRGLVYEISLOEALAIYGNSSPAAMTTRVVDGFGGKYYTP 378
DB 12983 GTSILGRSKAYIND--QDQOLV-----PAGVVGELVAVGDG----- 13016
QY 379 LTRGVDCPYLATVDMHFLLESQAARTIRDAFCVFEDONGLPLRRHNSDLXSHFGSLAE 438
DB 13017 LANGYDPSLNTGRFTHITIDGQVQAYR-----TGDRAVKRPDRDYQIEFFRLDQ 13067
```

```

QY 439 TLVYVRSMSTLLANDYWDVYFHPSCAIEIFATGYISSAFELGATGCKYGNV----- 492
DB 13068 QIKIRGHR-----TEPA-EVEQALLSDSSINDAVVYSAONKEGLELVGYITT 13113
QY 493 -----SEHTLGTVHTHSAHF-----VLDVAGLE--NMVNAEDVVFPMVAPV 534
DB 13114 QAAQSVDEKESNKNVOEWEAHPDSTAYANIGIDRDALGQFSLMTSMTYDOSLIPR---- 13169
QY 535 SPEHQLORLOVT--RKLLEMEQAAFL--VGSATPRYLYLASHNSKMGHPRCGRITOMLSFA 592
DB 13170 --EEMOEVLNDMTMSRLDNDQPGKVKLEIGTGTGVLF-----NLGVEG--LQSYA 13216
QY 593 GEPLPQSSMARGFSWERYQLAIVQRKEEPPSSSVFQNDPMAPTYDFSFINNETIA- 651
DB 13217 G--LEPSRSVT--AW-----VNKALETFFSLAGSARVHGTAEEDISSIDGLSLDLVI 13265
QY 652 -----GKDLVAMVTAGFLIHPAEDIPRTVTVGNGVGF--LRPY-----NF----- 691
DB 13266 NSVAQYFPRSREYLAELTANLIRLPGVKRI-----FFGDMKRYATATKDLVARAV 13314
QY 692 -----FDEDPSTYSADSIYFRGDQDAGACEVNPPLACLPOAAACAPDL 733
DB 13315 HTLGSNASKAMVROQVAKLEDEDEELLVDPAPAFSLSDQFPDEIKHYELLPKRMAATNEL 13374
QY 734 PAFSHGGPSH 743
DB 13375 SSRYAAVTH 13384

RESULT 10
PCT-US94-03437-8
? Sequence 8, Application PC/TUS9403437
? GENERAL INFORMATION:
? APPLICANT:
? TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT-
? TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASES
? NUMBER OF SEQUENCES: 124
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/03437
? FILING DATE:
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 485 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Leuconostoc dextranicum
? STRAIN: ATCC 19255
? PCT-US94-03437-8

Query Match      2.3%; Score 93.5; DB 5; Length 485;
Best Local Similarity 20.8%; Pred. NO. 0.95;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

QY 387 YLATYVDMHF-----LLESQAARTIRDAFCVFEDONGLPLRRHNSDLXSHYFGLA 437
DB 30 YKGYLDGHEFAIVGTARQGSDDERKQLVDSIKDFEDQA-----QAENFLAHFETRA 83
QY 438 ETVLVVNASMTL-----LNTDYWDVYFHPSCAIEIFAT--GYISSAPLFGATGK 487
DB 84 HDVTDAASYGILKSAIEAATKFDIDGNRIIFYMS--VAPRFGTIAKYLKSEGLAETG- 140
```


COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/445,464C
APPLICATION NUMBER: US/08/445,464C
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/044,857
FILING DATE: 08-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Ruzsala, Lois K.
REGISTRATION NUMBER: 39,074
REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 267-5364
TELEFAX: (847) 267-5376
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-445-464C-8

Query Match 2.38; Score 93.5; DB 3; Length 486;
Best Local Similarity 20.88; Pred. No. 0.96;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

QY 387 YLATYVDMHF-----LLESQAPRTIRDAFCVFQONOGPLRRHNSDLXSHYFGGLA 437
DB 31 YKKGYLQEHFAIVGTARQQLSDDEFKQLVRSIDFTEDDA-----QAEAFIHFSTR 84
QY 438 ETVLVVSMSTL-----LNTDYVMDVTFHPSGAIEIRFYAT--GYISSAFLEGATGK 487
DB 85 HDVTDAASYGILKSAIEAATKFDIDGNRIFYMS--VAPRFPGTIAKYLKSEGGLAETG- 141
QY 488 YGNQVSEHTLGTVHTSAHFKVDLVAGLENMVM-----AEDVETVPMAY----- 532
DB 142 YNRLMIKPEFGTSYATAEELQSDLENAFDDQLEFRIDHYLKEKVNQNIALRFGNPIFDA 201
QY 533 PMSPEHOLQLOVY--RKLEMEBOAFLVGSATPRYLYLASNHSNKGHPRGYRIOMLSF 591
DB 202 AMNKDY--IKNQVTLAEVLGEERAGY--DTTGALLDMIONHT-----MQIYGM 248
QY 592 AGEPLQ--NSSMARGESWERYQLAVTQKKEEPSVSSVFQNDPMAPTVDSDFI---- 645
DB 249 LAMEKPEFNDKDIRAKNAF--ALKIYNEEYVKYFVRAQYGA--GDTADYKPYLEAD 306
QY 646 -----NNETIAGK--DLVAVTAGFLHPIHADIPTVTYGN-----GVGFPLR 687
DB 307 VPADSKNNTFIAGELQFLPRM-----EGVPEYVSGRLAKAKOTRVDIVFKAG 355
QY 688 PYNFDEDPFSYSADSIYFRGDDAGACEV 717
DB 356 TFMGSEDAQESVLSITIT---DPKGAIEL 382

RESULT 15
US-08-044-857D-8
Sequence 8, Application US/08044857D
Patent No. 6455288
GENERAL INFORMATION:
APPLICANT: Jakobovits, Edward B.
Silen, Joy L.
Levy, Mark J.
Goodman, Thomas C.
Becker, Martin

Caldwell, Robert M.
Bolt, Richard R.
Barnett, Christopher C.
TITLE OF INVENTION: Homogeneous Immunoassays Using Mutant
Glucose-6-Phosphate Dehydrogenases
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dade Behring Inc.
STREET: 1717 Deerfield Road
CITY: Deerfield
STATE: Illinois
COUNTRY: USA
ZIP: 60015

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/044,857D
APPLICATION NUMBER: US/08/044,857D
FILING DATE: 08-APR-1993
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ruzsala, Lois K.
REGISTRATION NUMBER: 39,074
REFERENCE/DOCKET NUMBER: BEH-7261
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 267-5364
TELEFAX: (847) 267-5376
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Leuconostoc dextranlicum
STRAIN: ATCC 19255
SEQUENCE DESCRIPTION: SEQ ID NO: 8
US-08-044-857D-8

Query Match 2.38; Score 93.5; DB 4; Length 486;
Best Local Similarity 20.88; Pred. No. 0.96;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

QY 387 YLATYVDMHF-----LLESQAPRTIRDAFCVFQONOGPLRRHNSDLXSHYFGGLA 437
DB 31 YKKGYLQEHFAIVGTARQQLSDDEFKQLVRSIDFTEDDA-----QAEAFIHFSTR 84
QY 438 ETVLVVSMSTL-----LNTDYVMDVTFHPSGAIEIRFYAT--GYISSAFLEGATGK 487
DB 85 HDVTDAASYGILKSAIEAATKFDIDGNRIFYMS--VAPRFPGTIAKYLKSEGGLAETG- 141
QY 488 YGNQVSEHTLGTVHTSAHFKVDLVAGLENMVM-----AEDVETVPMAY----- 532
DB 142 YNRLMIKPEFGTSYATAEELQSDLENAFDDQLEFRIDHYLKEKVNQNIALRFGNPIFDA 201
QY 533 PMSPEHOLQLOVY--RKLEMEBOAFLVGSATPRYLYLASNHSNKGHPRGYRIOMLSF 591
DB 202 AMNKDY--IKNQVTLAEVLGEERAGY--DTTGALLDMIONHT-----MQIYGM 248
QY 592 AGEPLQ--NSSMARGESWERYQLAVTQKKEEPSVSSVFQNDPMAPTVDSDFI---- 645
DB 249 LAMEKPEFNDKDIRAKNAF--ALKIYNEEYVKYFVRAQYGA--GDTADYKPYLEAD 306
QY 646 -----NNETIAGK--DLVAVTAGFLHPIHADIPTVTYGN-----GVGFPLR 687
DB 307 VPADSKNNTFIAGELQFLPRM-----EGVPEYVSGRLAKAKOTRVDIVFKAG 355
QY 688 PYNFDEDPFSYSADSIYFRGDDAGACEV 717

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 12:04:34 ; Search time 25 Seconds
(without alignments)
2951.379 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763

Perfect score: 4010
Sequence: 1 LVCVLLVGRGSGDGEPSQLP.....QAAACAPDLPAFSGHGFSSHN 744

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PTCT_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4010	100.0	763	US-10-081-408-2	Sequence 2, Appl1
2	3967	98.9	998	US-10-081-408-20	Sequence 20, Appl1
3	2489.5	62.1	729	US-09-919-497-51	Sequence 51, Appl1
4	106	2.6	617	US-10-005-530-58	Sequence 58, Appl1
5	104	2.6	980	US-09-908-193-22	Sequence 22, Appl1
6	103	2.6	2224	US-10-115-563-14	Sequence 14, Appl1
7	102	2.5	1194	US-09-738-626-5832	Sequence 5832, Ap
8	101	2.5	617	US-10-005-530-46	Sequence 46, Appl1
9	99	2.5	944	US-10-213-990-27	Sequence 27, Appl1
10	95.5	2.4	581	US-09-738-626-4914	Sequence 4914, Ap
11	95	2.4	617	US-10-005-530-57	Sequence 57, Appl1
12	93.5	2.3	878	US-09-912-020-364	Sequence 364, App
13	93	2.3	601	US-09-925-301-844	Sequence 844, App
14	92.5	2.3	635	US-10-099-895-33	Sequence 33, Appl1
15	92	2.3	904	US-09-712-363-247	Sequence 247, Appl1
16	91.5	2.3	994	US-10-004-551-16	Sequence 16, Appl1
17	91	2.3	390	US-09-738-626-6558	Sequence 6558, Ap
18	91	2.3	1247	US-09-808-193-2	Sequence 2, Appl1
19	91	2.3	2507	US-09-819-104A-2	Sequence 2, Appl1

20	90.5	2.3	388	US-09-361-630-1	Sequence 1, Appl1
21	90	2.2	589	US-09-740-041-2	Sequence 2, Appl1
22	90	2.2	617	US-10-005-530-56	Sequence 56, Appl1
23	89.5	2.2	630	US-10-101-464A-71	Sequence 71, Appl1
24	89.5	2.2	911	US-09-855-754-4	Sequence 4, Appl1
25	89.5	2.2	911	US-10-227-353-4	Sequence 4, Appl1
26	89.5	2.2	953	US-09-845-583-4	Sequence 4, Appl1
27	89.5	2.2	3695	US-10-037-182-2	Sequence 2, Appl1
28	89	2.2	365	US-09-925-1007	Sequence 1007, Ap
29	89	2.2	365	US-09-925-299-1007	Sequence 1007, Ap
30	89	2.2	1430	US-10-087-993-34	Sequence 34, Appl1
31	88.5	2.2	583	US-09-942-185-2	Sequence 2, Appl1
32	88.5	2.2	616	US-09-925-300-1519	Sequence 1519, Ap
33	88.5	2.2	882	US-10-174-590-574	Sequence 574, App
34	88.5	2.2	882	US-10-176-758-574	Sequence 574, App
35	88.5	2.2	882	US-10-175-737-574	Sequence 574, App
36	88.5	2.2	882	US-10-173-706-574	Sequence 574, App
37	88.5	2.2	882	US-10-175-738-574	Sequence 574, App
38	88.5	2.2	882	US-10-175-752-574	Sequence 574, App
39	88.5	2.2	882	US-10-176-482-574	Sequence 574, App
40	88.5	2.2	882	US-10-176-757-574	Sequence 574, App
41	88.5	2.2	882	US-10-176-913-574	Sequence 574, App
42	88.5	2.2	882	US-10-180-552-574	Sequence 574, App
43	88.5	2.2	882	US-10-180-557-574	Sequence 574, App
44	88.5	2.2	882	US-10-173-700-574	Sequence 574, App
45	88.5	2.2	882	US-10-174-572-574	Sequence 574, App

ALIGNMENTS

RESULT 1
US-10-081-408-2
Sequence 2, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abirams, n, Lars
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: US 60/272,247
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 763
TYPE: PRT
ORGANISM: Homo sapiens
US-10-081-408-2

Query Match 100.0%; Score 4010; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 744; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	LVCVLLVGRGSGDGEPSQLPHCPSPVSPAPWTHPQOSQFALSRRELTAVNRFTLORL	60
DB	20	LVCVLLVGRGSGDGEPSQLPHCPSPVSPAPWTHPQOSQFALSRRELTAVNRFTLORL	79
QY	61	GPGLVDAARPDNCFVSVELDLPKKAALALHLDGSPPARREALAIVFGQPOPNN5	120
DB	80	GPGLVDAARPDNCFVSVELDLPKKAALALHLDGSPPARREALAIVFGQPOPNN5	139
QY	121	ELTVGGLPHPSYRDVTVRRHGPPLPYHRRPVLFQEYLDIDQIFNRELPOASGLHHCC	180
DB	140	ELTVGGLPHPSYRDVTVRRHGPPLPYHRRPVLFQEYLDIDQIFNRELPOASGLHHCC	199
QY	181	FYKHRGRNLTMTATARGLOSGRATWFGLYXNISGAFLLHVGLELLVNHKALDPARW	240
DB	200	FYKHRGRNLTMTATARGLOSGRATWFGLYXNISGAFLLHVGLELLVNHKALDPARW	259

Qy	241	TIQAFYVQGRYVDSLAQLEAENQFAGLVNVVLIIPDNGTGGMSLKSVPVPGAPAPLQAFPO	300
Db	260	TIQAFYVQGRYVDSLAQLEAENQFAGLVNVVLIIPDNGTGGMSLKSVPVPGAPAPLQAFPO	319
Qy	301	GPRFSVQGSRAVSSLMFTFSFGGLGAFSGPRIDVDFQGERLYEETISLOALATYGNSPPAA	360
Db	320	GPRFSVQGSRAVSSLMFTFSFGGLGAFSGPRIDVDFQGERLYEETISLOALATYGNSPPAA	379
Qy	361	MTTRYVDGFGMGKRYTTPPLTRGVDCPYLATYVDMHFLLESQAPKTIIRDACVFEEQNGLP	420
Db	380	MTTRYVDGFGMGKRYTTPPLTRGVDCPYLATYVDMHFLLESQAPKTIIRDACVFEEQNGLP	439
Qy	421	LRRHSDLSYHFEGGLAETVLYVRSMSSTLLWYWDYVMDYVFPSCAIEIRFATGYISAF	480
Db	440	LRRHSDLSYHFEGGLAETVLYVRSMSSTLLWYWDYVMDYVFPSCAIEIRFATGYISAF	499
Qy	481	LEGATGKGNQVNSEHTTGCTVHTHSAHFYVDLDVAGLEKMWVAEDVAFVPMVAPMSPEQL	540
Db	500	LEGATGKGNQVNSEHTTGCTVHTHSAHFYVDLDVAGLEKMWVAEDVAFVPMVAPMSPEQL	559
Qy	541	QRLQVTRKLLMEBOAAELVGSATPRYLTYLASNHSNKGHPRGYRIOMLSEAGEPLPONS	600
Db	560	QRLQVTRKLLMEBOAAELVGSATPRYLTYLASNHSNKGHPRGYRIOMLSEAGEPLPONS	619
Qy	601	SMARGFSMERQOLAVTQKKEEPPSSSYFNONDPAPVYDSDFIINNTTIGKDLVAVYT	660
Db	620	SMARGFSMERQOLAVTQKKEEPPSSSYFNONDPAPVYDSDFIINNTTIGKDLVAVYT	679
Qy	661	AGFLIIPHAEDIPMTVTYNGNGVGFLLRPYNFDEDEPSFYSDSYFRGDQAGACEVNP	720
Db	680	AGFLIIPHAEDIPMTVTYNGNGVGFLLRPYNFDEDEPSFYSDSYFRGDQAGACEVNP	739
Qy	721	ACLPOAAACAPDLPAFSHGFSHN	744
Db	740	ACLPOAAACAPDLPAFSHGFSHN	763
RESULT 2			
US-10-081-408-20			
Sequence 20, Application US/10081408			
Patent No. US20020160482A1			
GENERAL INFORMATION:			
APPLICANT: Abiatus, n, Lars			
APPLICANT: Nilsson, Joakim			
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION			
FILE REFERENCE: 13425-053001			
CURRENT APPLICATION NUMBER: US/10/081,408			
CURRENT FILING DATE: 2002-02-21			
PRIOR APPLICATION NUMBER: SE 0100625-3			
PRIOR FILING DATE: 2001-02-23			
PRIOR APPLICATION NUMBER: US 60/272,247			
PRIOR FILING DATE: 2001-02-28			
NUMBER OF SEQ ID NOS: 20			
SOFTWARE: FastSeq for Windows Version 4.0			
SEQ ID NO 20			
LENGTH: 998			
TYPE: PRT			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Recombinant construct			
US-10-081-408-20			

	Query Match	98.9%	Score 3967	DB 9	Length 998
	Best Local Similarity	99.9%	Pred. No. 0		
	Matches 735	Conservative 1	Mismatches 0	Indels 0	Gaps 0
Qy	9	RGDDGGEPSQLPHCVSPSAQPMTHPGGSGFADLSRELTAVKRFQRLGPGIVDA	68		
	:				
Db	263	QGDDGGEPSQLPHCVSPSAQPMTHPGGSQLFADLSRELTAVKRFQRLGPGIVDA	322		
Qy	69	QARPSDNCVSVLEQLPPKAAALAHLDKSGPPAREALAIYFEGRQPOENVSELVYGPLP	128		

Dd	323	QARSNDNCVSVLEQLPRKAAALAHDRGSPPARRELAIVFFGRQRPQPVNSLVVGPRL	382
Qy	129	HPSYMRDVYERHGGPLPYHRRPVLEFOEYLDIDOMITNRELPOASGLLHHCCFYKHKGRN	188
Dd	383	HPSYMRDVYERHGGPLPYHRRPVLEFOEYLDIDOMITNRELPOASGLLHHCCFYKHKGRN	442
Qy	189	LVMTTTPARGLQSDRATWMLGXYNISGAPFELHHVLELLVHNHKAIDPARMTIOGVFO	248
Dd	443	LVMTTTPARGLQSDRATWMLGXYNISGAPFELHHVLELLVHNHKAIDPARMTIOGVFO	502
Qy	249	GRYDYSLAQLEAQEAGLVNWNVLIPDNGTSGSMSLKSPVPGAPRPLQFVPOGPRFSVQ	308
Dd	503	GRYDYSLAQLEAQEAGLVNWNVLIPDNGTSGSMSLKSPVPGAPRPLQFVPOGPRFSVQ	562
Qy	309	SRVASSLMTFSFGIAGTSPRIEDVBRQGERLYEISLOALAIYGGNSPAAMTTRYVQ	368
Dd	563	SRVASSLMTFSFGIAGTSPRIEDVBRQGERLYEISLOALAIYGGNSPAAMTTRYVQ	622
Qy	369	GFGMGKTYTPTLRGVDPDYLATVYDMHFLESOAPKTIIPAFCVEFQNOGLPLRRHNSDL	428
Dd	623	GFGMGKTYTPTLRGVDPDYLATVYDMHFLESOAPKTIIPAFCVEFQNOGLPLRRHNSDL	682
Qy	429	YSHFPGGLAEFVLYVVRKSKSTLANTDYWDYVFNHSGAIEIRFATGTYSSAFIFGATGX	488
Dd	683	YSHFPGGLAEFVLYVVRKSKSTLANTDYWDYVFNHSGAIEIRFATGTYSSAFIFGATGX	742
Qy	489	GNQVSEHTLGTVHTHSAFKVDLVAGLENNVMAEDVAFVPMVPMSPHQRLQVTRK	548
Dd	743	GNQVSEHTLGTVHTHSAFKVDLVAGLENNVMAEDVAFVPMVPMSPHQRLQVTRK	802
Qy	549	LLEMEQAPLVGSATPRYLYLASNSHKNKGHPRGYTIOMLSPAGEPLPONSSMARCFSW	608
Dd	803	LLEMEQAPLVGSATPRYLYLASNSHKNKGHPRGYTIOMLSPAGEPLPONSSMARCFSW	862
Qy	609	ERYOLATVORKEEPPSSSVFNQNDPMAPRYTDFSDFINNETIAGKDLVAVWTAAGFLIIP	668
Dd	863	ERYOLATVORKEEPPSSSVFNQNDPMAPRYTDFSDFINNETIAGKDLVAVWTAAGFLIIP	922
Qy	669	AEDIPNTVTVNGVGFELRPYNFEDEDEPSFYSAISYIFRGDQDAGACEVNPPLACLPOAAA	728
Dd	923	AEDIPNTVTVNGVGFELRPYNFEDEDEPSFYSAISYIFRGDQDAGACEVNPPLACLPOAAA	982
Qy	729	CAPDLPATFSHGGSFN 744	
Dd	983	CAPDLPATFSHGGSFN 998	

```

RESULT 3
US-09-919-497-51
: Sequence 51, Application US/09919497
: Patent No. US2002010662A1
: GENERAL INFORMATION:
: APPLICANT: Mutter, George L.
: TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
: FILE REFERENCE: B0801/7225
: CURRENT APPLICATION NUMBER: US/09/919,497
: CURRENT FILING DATE: 2001-07-31
: PRIOR APPLICATION NUMBER: US 60/221,735
: PRIOR FILING DATE: 2000-07-31
: NUMBER OF SEQ ID NOS: 100
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 51
: LENGTH: 729
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-919-497-51

```

Query Match	62.1%	Score 2489.5;	DB 10;	Length 729;
Best Local Similarity	63.7%	Pred. No. 4.2e-211;		
Matches 473; Conservative	86;	Mismatches 149;	Indels 35;	Gaps 4;
QY	1	LVCVLLVGRGSGGEPESQLPHCPSPVSPSAQPTHTGQSQLFADLSREELTAVMRFLLTORL	60	

```
Db 20 LAYVLLTSPGGS---SQPHCPVSHRAQMPHPGSQLFADLSFEELTAVNRILTQRL 75
Oy 61 GGLVDAQAARPSDNCVFSEVLEQLPPKAAALAHLDGSPPPAREALATVFGQPOPNIS 120
Db 76 GGLVDAQAQAQSDNCIFSEVLEQLPPKAAALAHLDGSPPPAREALATVFGQPOPNIS 135
Oy 121 ELVVGRLPHPSYMRDVTVERHGGPLPYHRRPYLFOEYLDIDOMIFNRELPOASGLLHCC 180
Db 136 ELVVGRLPHPSYMRDVTVERHGGPLPYHRRPYLRAEFTQMRRLKQVLEPKAFILSSR- 194
Oy 181 FKHHRNRLVMTAPRGVQSGDRATWFGLYTISGAGFLHVGLELLVNHKALDPARM 240
Db 195 -FNYNSTLAAVAHAPRGRLRSRRTTWGLYHNSGVGLFHPVGLLELDHDLDPARM 253
Oy 241 TIOKVYOGRRYDLSLAQAOFAGLVNVLLIPDNGTGSWSLKSVPVPPAPLPFOYPO 300
Db 254 TVQGVYTLGHYVADLQGLEREFSGRLVRYVPLPPNGASLSRSNSFGPLPPLPFSQ 313
Oy 301 GPRFSVQSGRVAASLTWTFSGLCAFGSPRIFDVRFQGERLYVEISLQALATYGSNSPAA 360
Db 314 GSGYSVQGNLVVSSLSWTFSGHGVFSGRLFDVRFQGERLAYEVSVQECVSIYGAOSPRT 373
Oy 361 MTRIVYDGGFGMGKITTPLTRGVDCRYLATYVDMHFLBESQAKTRDAFCVEQNGLP 420
Db 374 MLTRYLDSFGLGRNSRGVGVDCRYQATMDIHLVKGAVQLLPAGAVCFEEAOLP 433
Oy 421 LRRHSDLSHYEGLAETVLYVRSMTLLNDYVMDTVFHPGSALEIRFATGYISSAF 480
Db 434 LRHHNYLQNHFEYGLASLALYVRSVSVGNVYIMDEVLYPRGALGRHAQYINTAF 493
Oy 481 LEGATG--KYGNVSEHTLGTVTHSAHEKVDIDVAGLENMVAEDMVPVMAVPSPEH 538
Db 494 LKQGEGLLFGNRYGVRGLGTVTHAFHEKLDIDVAGLKMVVAEDVFEKPAAPNPEH 553
Oy 539 QLORLVTRKLEMEBOAFLVGSATPRYLYLASNSNKKGHPRGYRIQMLSTAGEPLPQ 598
Db 554 WLORPOLTRVYLGKEDITAFSLGSPLPRIYLYLASNOTNAMGHORG----- 598
Oy 599 NSSMAGFSEWERYQALAVTORKEEPESSSVFNQNDPAPVDESDFINNETIAGKDLVAM 658
Db 599 -----YQLVYQKKEESQSSSIYHQNDLWTPTVTAFDFINNETLIGEDLVAM 646
Oy 659 VTAGFLHIAHEDIPNTVYVNGVGFPLRYNFDDEPSTYSADSIYFRDQAGACEVN 718
Db 647 VTSFLHIAHEDIPNTVYVNGVGFPLRYNFDDEPSTYSADSIYFRDQAGACEVN 706
Oy 719 PLACLPQAAACAPDLPAFSHGF 741
Db 707 PVACLPLDLAACVPDLPEFSTHGF 729
```

```
RESULT 4
US-10-005-530-58
; Sequence 58, Application US/10005530
; Publication No. US20030026795A1
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
; FILE REFERENCE: 11899.0022.DVUS01 (MOBT:022--2)
; CURRENT APPLICATION NUMBER: US/10/005,530
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/063,733
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/044,504
; PRIOR FILING DATE: 1997-04-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 617
; TYPE: PRT
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Polypeptide
; NAME/KEY: MISC_FEATURE
; LOCATION: (74)..(79)
; OTHER INFORMATION: Xaa = Unknown
US-10-005-530-58

Query Match          2.6%; Score 106; DB 9; Length 617;
Best Local Similarity 22.8%; Pred. No. 1.3;
Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;

Oy 44 LSREELTAVNRILTQRLG--PGLVDAQAAR-----PSDNCVFSEVLEQLPPKAA 90
Db 18 LIREKVAKEILNTERLGEVPGIPRNEGRFLGSGYSHDNLPSDPLYSISK-----PGCGG 73
Oy 91 LAHLDSGPPPARREALATVFFGQPOPNVSELYVG-----PLPHPSYMRDVTVERHGGP 144
Db 74 XXXXXGGGPPRK--VCIVGAG-----VSGLYIAMILDOLKIPNLTYDIFESSSTTGR 125
Oy 145 LPYHRRPVLFOEYLDIDOMIFNRELPOASGLHCCFFYKHGRNLTMTAPRGLOSGDR 204
Db 126 LYTHHTDAKHQDYDGNARY-----PDIPSMKRTNLFKRTGMPLI-----KYIYDGEN 175
Oy 205 ATWFGLYYINSGAGFLHVGLELVN-----HKALDPARMTOQVYOGRRYDLSLA 256
Db 176 TP--QLYNN--HFFAKGVSDPYNVSVANGTVPDDVDVSGEKLQOAF--GYREKLA 227
Oy 257 QLEAOFEAGLVNVLLIPDNGTGSWSLKSVPVPPAPLPFOYQGRFSVQSGRVAASLSL 316
Db 228 E--DFDKGDELMLVDMDMT--REYLKRGGRGGAEPKDF-----FALQMETON---- 273
Oy 317 TFSFGLGAFSGPRIFDVRFOGERLYVEISLQ-----EALATYGSNS 357
Db 274 -----TGTNLFPQAF--SESVIDSFDFDNTKPEWCIEGCTS 309
```

```
RESULT 5
US-09-908-193-22
; Sequence 22, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, NURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-908-193-22
```

Query Match 2.6%; Score 104; DB 9; Length 980;
Best Local Similarity 20.0%; Pred. No. 3.9;
Matches 13; Conservative 69; Mismatches 204; Indels 262; Gaps 37;

```
214 ISGAGFLHVGLELVNKAALDPA---WTICKVYQGRYDSLAQLE-AQEEAG-LV 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 ISGKYLWREVGAEENGRRLPGRGDQAMDVGPV---RLKKKVKQYELTQLVPGRL 443
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 NVYLIP---DNGGSGMSLSKSPVPGAPPLQFYPOGPRRSYOG-SRVASSLTFTSGL 322
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
444 EVKLVAENKHEGDAVAWKGKTEKAPADMDIQGPPLPAHVHAESSTVSIM-LRWKK 502
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323 GAFSGPRI-FDYAF-----QGERLYEISLQELAIYGNSPAMATRY----- 365
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
503 PDFETVAIVNTVAFSPMGLNANSLVYTISSGEDILI-GGLKP---FTKFEFAVQSHGV 558
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 -VDCGFPM-----GKTYPLTRGVDCPYLATVYDMHFLLESQAPKTIADACVEQN 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
559 DMDGPFSGSVYERSTLPDRSPSTDLRLSPLTSTVLHMCPTPEPNCIEIYELVSSN 618
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 QGLPLR-----HHSDLYSHY-----FGSLATVLYVRMS 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 HTQPEHMTLLTTGQNTFSAVHGLESDTRIFFKMGARTEVGPGFSLQDVITLQEKLS 678
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
448 TLNLYDVMDTVFHPGSAIEIRFYATGYISSAFL-----FG 483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
679 DSLMHNV-----TGIIYVCLGLCLLACMCAGLRSPHRESLELS 721
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 ATGKYGQV--SEHTLGTGTRHSAHFVVDLDVAGLENNVMAEDVFPMAVPSGPEHQL 541
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
722 STAPGPALYSRARLGPSPPPAAH-----ELESLVH-----PHODWSPPPS-- 764
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
542 RLQYTRKLEMEQAAF--LVGSATPRLYLASNSNK---WGHPRGVRIMSPAGSPL 596
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 -----DVEDRAEVHSLMGGVSE-----GRSHSKKISMAQPSG-----LSMAG-- 803
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
597 PONSMAKGFEMERYOL-----AVTQ-----RKEEEP 623
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804 -----SMAGCELPQAGPRPALTRALLPAGTGTLLQALVYDAIKNGKKRPP 853
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
624 SSSVFNONDPAPTVPSDFINNETIAGKDLVAMVTAAGFLIHPAEDI-----PNTV 676
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854 AC-----RNOVEAEVYHNSDF--SASNGNPDL-----HLODLEDEDP.LPPAP 894
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
677 TVNGVGFELRPVNFDEDPFSYASDIYFRG---DODAGACEY-----NPLACLPOA 727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
895 DLISGV-----DPC-----OGAAMLDRELGCCELAAFGPRLICLPEAA 934
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
728 ACA---PDL 733
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
935 SASCSYPLD 943
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 6
US-10-115-563-14
; Sequence 14, Application US/10115563
; Publication No. US20030008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffith, John H
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
; AND COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; Patent Counsel
; STREET: 10666 No. US20030008307A1ch Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/115,563
FILING DATE: 02-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/410,488
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: 449.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2224 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-10-115-563-14
Query Match 2.6%; Score 103; DB 9; Length 2224;
Best Local Similarity 20.7%; Pred. No. 16;
Matches 76; Conservative 46; Mismatches 123; Indels 122; Gaps 20;

```
150 RPLVFOEYLDIDOMIFNRELPOA--SGLLHHCDF-----YKHGRNLVTMTAPR 197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 KKIYREY-----EYFKKKEKQSTISGLGPTLYAEVGDIKVFKNKADK--PLSIHPQ 115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 GLQSGDRAVTFGLYYNIS-GAGFELHVGLELVNKAALDPA---WTIQK----- 244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 GIR-----YSKLSEGASVLDHTFPARKM--DDAVAGREYTYEWSISEDSGPTH 163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 -----VFQGRYDLSLAQLEAFEGLVNVVILPONGT---GG----- 279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
164 DPCLTHTIY-----SHENLIEDENSGLLGLPDLICKKGLTEGTOKTEQKQIVLLFAV 217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 -----SMLKSPVPGPAPPLQFYPOG--PRFSYQGSRVASSLTWTFSGGLAFSGPRI 332
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 FDESKSMQSSSL---MYIYNGVNGTMDPTICADHDIS--WHL---LQMSGPELFS 268
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 VRQGERLYEISLQELAIYGNSPAMATRYVDGFGMGKYYTP-LTRGVDCPYLATY 391
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 IHFMQVYLEQNHKHYSAITLVASATSTANMTVGEGKWIISLTPKHLQAG-----MQAY 323
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 VD-----WHFLLESQAPKTIIRDAFCVFQNGCLPLRRHSD 427
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 IDIKNCPKTRNLKKTREORRHMKRWEYFL--AAEVIWDYADVIRPANDKRYRSOHL 381
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 LYSHYFG 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 NFSNIG 388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7
US-09-738-626-5832
; Sequence 5832, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5832
LENGTH: 1194
TYPE: PRF
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5832

Query Match 2.5%; Score 102; DB 9; Length 1194;
Best Local Similarity 18.9%; Pred. No. 7.9;
Matches 112; Conservative 71; Mismatches 207; Indels 204; Gaps 29;

QY 195 APRGLOSGRATWFGLYY---NISGAFGLHVGLELVNHNKALDPAKMTIQKVFYQGRY 251
DB 80 APSPRFKEKRVHNGEPHQSDVSGGAYLHQ---MLAENNT---GLRNLDF---Y 126
QY 252 YDSLAQLEAFEGALVNVVLPDNGTGSWSLSPVPBPAPLQFPGPSPRSVQGSRV 311
DB 127 LSSMASTEGCL-----GKM-----PRND----- 144
QY 312 ASSLMFSGFLGAFSGPRIDVRFQGERLYEISLQALAL---IYGNSPAAMTRYVDG 368
DB 145 ADIAEHAEGIIATTCGSPSDVQTRLRIGQFDEALAAAMODIYGRDN---YFLEIMDH 201
QY 369 GFG---MGKTTPLTRNGVCPYIATYVDHNFLESQAPRTIRAPFCVFQDNOGLPLRRH 424
DB 202 GLDIETFRSELLEIGKRLNPLVTN-DCHVYLESQA--QAHBAMLCVGTGKTL---H 254
QY 425 HSDLSHYFGGLAETVYVNSMSTL-----LNYDYVMDTVFHPSCA 465
DB 255 DEDRFK---FGGTGYVVASAQMRALMDMPDGCNDTLMIAERQSTDEITEE--HSHDR 310
QY 466 IEIRFYATGYISSAFLEFGATGKXGNOVSEHTLGTVHTSAHFKYDLDVAGLENNVMAEDM 525
DB 311 MPADVPEGYTPPTWL-----HHEV---MAGLEDREFSGQV 343
QY 526 VFPVMAVPSPEHQLQYTRKLLMEEOAFLVGSATPRYLYLAS---NHSN---KW 578
DB 344 -----PEDIERAEYEISVIDMK-----GYPSYFLIVEIILKHARSIGIRV 384
QY 579 GHPRGYRIOMLSFAGEPLPQSSMARGFSEMERQYLAVTQKKEEPPSSSVENQDPMAP- 637
DB 385 GPGHGSAGALVAVALTITNIDPEHGLPERF-----LNPERPSAPD 427
QY 638 -TVDFSDFINNETIA-----GKDLVAV- TAGFLHIPHAEIDPNTYVNGVGFPLRPY 689
DB 428 IDIDFDRRREGEMIRYAADRWGEDKIAOVIFGTVKTKQALKDSARVQMG----- 477
QY 690 NFEFEDDSFTYSADSIYFRGDODAGACEVNPAC---LPQAAACAPDLPAFBSHG 740
DB 478 -----QPGYQYADRV-----IKELPAPIMAKDIPLSGITDPDHPRFNEAG 517

RESULT 8
US-10-005-530-46
Sequence 46; Application US/10005530
Publication No. US20030026795A1
GENERAL INFORMATION:
APPLICANT: Isaac, Barbara G.
APPLICANT: Greenplate, John T.
APPLICANT: Purcell, John P.

APPLICANT: Romano, Charles P.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
FILE REFERENCE: 11899,0022.DVUS01 (MOBT:022-2)
CURRENT APPLICATION NUMBER: US/10/005,530
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: 09/063,733
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/044,504
PRIOR FILING DATE: 1997-04-21
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 46
LENGTH: 617
TYPE: PRF
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Polypeptide
US-10-005-530-46

Query Match 2.5%; Score 101; DB 9; Length 617;
Best Local Similarity 23.3%; Pred. No. 3.6;
Matches 81; Conservative 39; Mismatches 138; Indels 90; Gaps 19;

QY 44 LSRBELTAVMRFLTORLG--PGLVDAQAAR-----PQONCVSEYELQPPKAA 90
DB 18 LIREKAKELNITLTERGVEVPIPPNVEGRFLGGYSHNDLPSPDLYSST-----KPL 71
QY 91 LAHLDRGSP--PPAREALATVFFGRQPPQVNSSELVVG-----PLPSPYMRDVTYERHG 142
DB 72 LKEAPRAEELPPRK--VCIVGAG-----VSGLYIAMILDDKIPRLTYDIFESSSRTG 123
QY 143 GPLPYRRPVLFOEYLDIDQMTFRELPOASGLHHCCKFKHRNRLVMTAPRGQSG 202
DB 124 GLTYHTFTDAKHDYDYGAMRY---PDIPSMKRTFNLRKRTGMPLI-----KYULDG 173
QY 203 DRATWFGLYNISGAFGLHVGLELVN-----HKALDPAKMTIQKVFYQGRYDS 254
DB 174 ENTP--OLYNN---HFAKGVSDPYVAVANGSTVDDVDVDSGEKLOAF--GYKKEK 225
QY 255 LAQLEAFEGALVNVVLPDNGTGSWSLSPVPBPAPLQFPGPSPRSVQGSRVASS 314
DB 226 LAE---DFDKGFEDLMVDDMT--REYLRKGRGKGEAPKYDF-----FAIQMETON- 273
QY 315 LMTFSFGLGAFSGPRIDVRFQGERLYEISLQ-----EALATIGNS 357
DB 274 -----TGTNLFDQAF--SESVYDSFDFDNPTRPEWYCIGGTS 309

RESULT 9
US-10-213-990-27
Sequence 27; Application US/10213990
Publication No. US20030082595A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Bussey, Howard
APPLICANT: Scorms, Reg
APPLICANT: Roemer, Terry
TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUNIGATUS ENCODING INDUSTRIAL
FILE REFERENCE: 10182-019-999
CURRENT APPLICATION NUMBER: US/10/213,990
CURRENT FILING DATE: 2002-08-05
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 944
TYPE: PRF
ORGANISM: Aspergillus
US-10-213-990-27

Query Match 2.5%; Score 99; DB 9; Length 944;
Best Local Similarity 20.2%; Pred. No. 10;
Matches 122; Conservative 62; Mismatches 243; Indels 176; Gaps 29;

LOCATION: (74)..(79)
OTHER INFORMATION: Xaa = Unknown
US-10-005-530-57

Query Match 2.4%; Score 95; DB 9; Length 617;
Best Local Similarity 22.5%; Pred. No. 12;
Matches 78; Conservative 40; Mismatches 142; Indels 86; Gaps 18;

QY 44 LSRELTAVMRRLTORLG--PGLVDAQAR-----PSDNCVFSEVLEQLPRAAA 90
DB 18 LREKAVAKELNITLRLGVPGLPPNEGRLGGVSHNDLPSDPLYSI-----GGSGG 73
QY 91 LAHLDRGSPPARREALATYFFGRQOPANSELVVG-----PLPHSYMRDVTVEHAGP 144
DB 74 XXXXXXGGGPPRRK--VCYIAG-----VSGLYIAMILDCLKIPNLTLYIFESSSRHTGR 125
QY 145 LPHYRRPVLFOEYLDIDQIFNRELPOAGSLHHCFCYKRGHNLVTMTAPRLQSGDR 204
DB 126 LYTHHTDAKHXYDYGAMRY-----PDIPSMKRTFNLFRITGMPL-----KYLDGEN 175
QY 205 ATWFGLYNISGAGFLHVGLELVN-----HKALDPARMTIQKVPYOGRTYDSIA 256
DB 176 TP--QLYNN---HFEKAVCSDPYVWSVANGTVPDDVYDVGKLOQAF--GYEKLA 227
QY 257 QLEAOFEGALVNVVLIIDNGTGSWSLKSPPPGAPPLQFYPOGRFSVGSRAVSSLM 316
DB 228 E--DDKGFDELMVDMTT--REYLKRGGRGEAPKYDF-----FALQMETON--- 273
QY 317 TFSFGIAGFSGPRIFDVRFOGERLYEISLQ-----EALAIYGSNS 357
DB 274 -----TGTLNLFQDAF--SESVIDSPDFDNTKPEWCIEGSGTS 309

RESULT 12

US-09-912-020-364
Sequence 364, Application US/09912020
Patent No. US20020045592A1
GENERAL INFORMATION:
APPLICANT: Zyskind, Judith
APPLICANT: Ohlsen, Karl L.
APPLICANT: Trawick, John
APPLICANT: Forsyth, R. Allyn
APPLICANT: Froelich, Jamie M.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
FILE REFERENCE: ELITRA 001DVI
CURRENT APPLICATION NUMBER: US/09/912,020
CURRENT FILING DATE: 2001-07-23
PRIOR APPLICATION NUMBER: 09/492,709
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: 60/117,405
PRIOR FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 485
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 364
LENGTH: 878
TYPE: PRF
ORGANISM: E. Coli
US-09-912-020-364

Query Match 2.3%; Score 93.5; DB 10; Length 878;
Best Local Similarity 20.2%; Pred. No. 28;
Matches 108; Conservative 61; Mismatches 190; Indels 175; Gaps 27;

QY 157 YLDIDQIFNRELPOA-----SGLHHCFCYKRGHNLVTMT--- 193
DB 148 HLDVGQORLNLITPOAFMSNRARGYIPPELWDPCINAGLLN---YNSGNSVQNRIGN 203
QY 194 --TAPRLQSG-----DRATWFGLYNN---ISGAGFLHVGLELVNKKALDP- 237

DB 204 SHYATLNQSGNLGAMRLRONTT---SYNSDRSSSGSKNMQHINTFL---ERDIIPL 257
QY 238 -ARMTIQKVPYOGRTYDSI-----AOLEAOFEGALVNVVLIIDNGTG- 278
DB 258 RSRRLTGDDYTOGDIFFDGFNFRGAQLASDNN-----MLPDSORGFARVHIGIARTAQ 310
QY 279 -----GWSLSKSPVPPGAPPLQFYPOGRFSVQ-----GS-----R 310
DB 311 VTIKONGDIYNTVPPGFTINDIYAAGNSGDLQVTTIADSGSIQIFTVPYSSVPLIOR 370
QY 311 VASSLMTFSFGIAGFSGPRIFDVRFOGERLYEISLQEALAIYGSNSPAAMTRTYDGGF 370
DB 371 EGHTRISITAGEYRSGNMQOEkTRFQSTLLH--GLPAMWTIYGGQ---LADRYARAF 425
QY 371 GNGKRTTPLLRGVDCPYLATYVDWHEFLBSQAPKTRDAFCVPEQNGQLPLRRHSDLYS 430
DB 426 GIGKNNAGLG-----ALSYDM-----TOANSTLPDD-----SQHDGGSVR---FLYN 464
QY 431 HVEGGLAEVLVVRSMSTLLNDYVDVFEH-----PSGALIRFPYATGYISSAFL 481
DB 465 KSLNESGTNIOLVGYRYSIGYFNEDITYSRNGINIEQDGVIOVKRFIDYVLA--- 522
QY 482 FGATGKYGNQVSEHTLG---TVHTSAR-----FKYDLVAGLEMMVMAEDMY 526
DB 523 YNKRGLQLQLTVNQ--LGRSTLYLSGSHQTYMCTSNVDEQFALNTA-----FED-- 572
QY 527 FVPMAVPSPEHOLQRLQVTRKLEMEBOAFLVGSATRIYLYLANSHSKNGH 580
DB 573 -----INWTLYSYL-----TKNAMQGRDOMALNVNIPSHVLRSDSKSOWRH 616

RESULT 13

US-09-925-301-844
Sequence 844, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 844
LENGTH: 601
TYPE: PRF
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (36)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (64)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (103)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (106)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (152)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (358)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (383)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-301-844

Query Match 2.3%; Score 93; DB 10; Length 601;
Best Local Similarity 22.3%; Pred. No. 18;
Matches 78; Conservative 32; Mismatches 112; Indels 128; Gaps 19;

QY 12 DGBEPQLPHC-----PSVPSA-QPWTP-----GOSQLADLSREE 48
DB 218 DTGTEALTPHIMNRLHATSRKSYRPSMEPMELSPFEDVACTEMSGDSGLDGS 277
QY 49 LTNMRFLTLGLP-GLVDAQAQP-----SDNCVSEVLEQLPPKAAALHLDRGSPPA 102
DB 278 QVS-SCPCSORSRSPDGLKKAEEPPKRPFGSSPLNAPCEGPP-----GSEPPR 326
QY 103 REALAIVFGRPOPNVSELVGLPPLPSYMRDVTVER--HGGLP-PYHR--RPLV--- 154
DB 327 RPPAPRHDGRKELPRQPLPPPI---GTERSOXTRGRGEPGRPSHREPPVQFGTX 383
QY 155 QEYLDI-----DOMIFNRE-----LPOASGLHHCFCYKRGRLVTM 192
DB 384 DKSDLLVYVGDLSLKAKEELTASVTEALPYSRDWELLPSAASA-----EPQSKNDISG 437
QY 193 TTPARGLQSDRATWFLYINISAGFFLHVGLLELVNKHAKDPAKWTQKVFYQGRY 252
DB 438 HCVEPPSSGQR-----LYPEVFY 456
QY 253 DSLAQLAEQFAGLVNVLIPDNG--TGGSMSL---KSP--VPPGPAP 294
DB 457 GSAPSSSQISGAMDSQLHPSGFGFPISLHPRSQPLYPGPAP 506

RESULT 14

US-10-099-895-33
Sequence 33, Application US/10099895
Patent No. US20020177166A1
GENERAL INFORMATION:
APPLICANT: BERNDT, Michael C
APPLICANT: STOMSKI, Frank C
APPLICANT: LOPEZ, Angel F
APPLICANT: GUTHRIE, Mark A
TITLE OF INVENTION: A BINDING MOTIF OF A RECEPTOR
FILE REFERENCE: 3991/OK379
CURRENT APPLICATION NUMBER: US/10/099,895
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: PCT/AU00/01118
PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 47
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33
LENGTH: 635
TYPE: PRT
ORGANISM: Homo sapiens
US-10-099-895-33

Query Match 2.3%; Score 92.5; DB 9; Length 635;
Best Local Similarity 21.9%; Pred. No. 21;
Matches 70; Conservative 32; Mismatches 107; Indels 111; Gaps 15;

QY 19 LPHCPSPSAQPMWTHRGOSOL-----FADLSREELTANMFLQRLGPGVDAQA 70
DB 132 LPAPPSIT-KAMGSOGELOJISWEPAPISDLRLYELRGPPDPKNSGTGPYI---QL 187
QY 71 RPSDNCVSEVLEQLPPKAAALHLDRGSP-----PPAREALAIYFGRQPOPNV 119
DB 188 IARETC--CPALGRPHASALDQSPCAQPTMPMDGPKQSPSEASALTRAEGG----- 239
QY 120 SELVGLPPLPSYMRDVTVER--HGGLP-PYHR--RPLV--- 154
DB 240 SCLISGLQPNNSYWLQLRSEPDGSLGSGWSWSLPTVYDLPGDAVALGLQCFLLDKNV 299
QY 170 -----POASGLLH-----C-----CYKRRGR 187
DB 300 TCOHODDHAASOGFFYHSRACCPDRDRIWENCEEEXTNPGLQTPQFSRCHFKSRND 359

QY 188 N-----LYTMTAPRGLOSGRATWFLYINISGAGFLHHV-----GLELVNKA 224
DB 360 SIHLIIVETTPA-----GTVHSYLSGSPFWIHQAVRLPPNLMHREISSGLE 407
QY 235 LD---PARMTQKVFYQGRY 251
DB 408 LEMHPPSSMAAQETCYOLRY 427

RESULT 15

US-09-712-363-247
Sequence 247, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
CURRENT FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 247
LENGTH: 904
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-247

Query Match 2.3%; Score 92; DB 9; Length 904;
Best Local Similarity 19.7%; Pred. No. 40;
Matches 132; Conservative 92; Mismatches 255; Indels 192; Gaps 37;

QY 155 QEYLDIDOMIFNRELPOAS-----GLHHCFCY-KHGRNLTMTAPRGLO 200
DB 8 KRFIDHVKAGHNEVPSASYILDPNLLFVNAGMVGQVFPFLQGRPPYFATASIDKICIR 67
QY 201 SGD-----RATWGLYINISGAGFLH---HGLELVNKA-----LDPAR-WTI 242
DB 68 TPIDDEVGITRNTFTEOMANGNSFGDYFRGALIELAMALLTNSLAAGYGLDPERIWT- 126
QY 243 QKVFYQGRY-----RYDSLAQLEAOFAGLVNVLIPDNGTGS--WSLSPVPPGPAPLOF 297
DB 127 -TYVFDDDEAVRLMOEYAGLPAP-----RIQRGMADNTWSKIPGCPSSSEI-Y 175
QY 298 YPOGPRFSVGSRYAS-----SLMTFSFGIAGSAGRIPDVRQGRRLYELISLQALAI 352
DB 176 YDRGPEFGAGPIYSBDRYLEVMNLVF-----MQRN--GEGTKEDYQI 219
QY 353 YGNSPAAMTTRYVDGFGNGK-----YTPPLTRGVDCPYLATIV----- 392
DB 220 LG-----PLPRKNIDTGMGVERIALYLVQDVHNYETDLR-----PVITFVARVARAYDV 270

```
OY 393 -----DHHEFLIESQAPKT-----IRDAFCVFEQNOGLPLRR-----HHSDLXSHYFG 434
Db 271 GNHEDDVARYRIADHSRTAAIILIGDVSPPGNDGRGYLRLLRRYIRSAKLLIGIDAIVG 330
OY 435 GLAEVVL-----VYRSMSTLLNDYVMDTFPHPSGAIEIRFYATGYISSAFLFGAT 485
Db 331 DLMATVRNAMGPSYPELVADFEERISRIVAEETAFNRRLTAGSSRLFEE-VASSTKKSAT 389
OY 486 GKYNQVSEHTLGTVH-THSAHFKVDLDVAGLENNVMAEDMVEVPM-----AVPWS 535
Db 390 VLSSGDAF-----TLHOTYGFPIELTEMAA-ETGLQYDEIGFRELMAEQRRRAKADAAA 443
OY 536 PEHOLQRLQVTRKLL-----EMEBQA-----AFLVGSATPRYLILASNHSNKMGH 580
Db 444 RKHAHADLSAYRELVDAGATEFTGFDELRSQARILGIFVDGKRPD-----VVAHGVAAGA 498
OY 581 PRGYRIOML-----SFAGEPLPONSMMARGFSWERYOLAVTORKEEPSSSSVFNQND 633
Db 499 GEGQREVELDRTPLIYASGGQIADGGTISGTGSSEARAATVDYOKIAKTL-----550
OY 634 PWA--PTVDFSDFINNETIAGKDLVAV--VTAGF--LHIPHA--EDI--PNTVTVG--N 680
Db 551 -VWHRVNVESGFEYEGDVIAAVDPGWRGATQGHSGTHMVAALROYLGPNAVQAGSLN 609
OY 681 GVGEFLRPYNF 691
Db 610 RPYGLRPFDFNW 620
```

Search completed: May 20, 2003, 12:09:46
Job time : 29 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:56:29 ; Search time 20 Seconds

(without alignments)
3576.200 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763

Perfect score: 4010

Sequence: 1 LVCVLLVGRGGDGEPSQLP.....QAAACAPDLPAFSGHGFSSHN 744

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4010	100.0	763	2 JC5234	amine oxidase (cop
2	3331	83.1	762	2 A54411	amine oxidase (cop
3	1398.5	34.9	746	2 S34656	amine oxidase (cop
4	1394	34.8	751	2 A54053	amine oxidase (cop
5	413	10.3	755	2 B41836	amine oxidase (fla
6	397.5	9.9	757	2 E64889	amine oxidase (cop
7	377	9.4	684	2 A56102	amine oxidase (cop
8	368.5	9.2	660	2 G90330	amine oxidase (cop
9	363	9.1	648	2 A48645	amine oxidase (cop
10	357.5	8.9	756	2 T48139	copper amine oxida
11	353.5	8.8	650	2 G71412	probable amine oxi
12	340	8.5	638	2 JC2139	phenylethylamine o
13	325.5	8.1	674	2 JC7251	amine oxidase (cop
14	324	8.1	687	2 T47403	amine oxidase-like
15	320	8.0	674	2 C44339	amine oxidase (cop
16	302.5	7.5	759	2 E84854	probable copper am
17	300	7.5	660	2 AH2234	copper amine oxida
18	295	7.4	692	2 S04963	amine oxidase (cop
19	247	6.2	794	2 T39171	probable peroxidom
20	246.5	6.1	300	2 T48138	copper amine oxida
21	229	5.7	587	2 S21139	amine oxidase (cop
22	214.5	5.3	670	2 S71320	amine oxidase (cop
23	126	3.1	862	2 T46289	hypothetical prote
24	125.5	3.1	460	2 T48137	copper amine oxida
25	125	3.1	757	2 C84120	subtilisin-type pr
26	111	2.8	1332	2 D82685	phosphoribosylform
27	107	2.7	435	2 A42672	choline kinase (EC
28	106.5	2.7	896	2 F96523	hypothetical prote
29	106	2.6	3938	2 T42761	basoon protein -

30	103	2.6	2224	1 KFHU5	coagulation factor
31	102.5	2.6	591	2 C48364	cytochrome-c oxida
32	101.5	2.5	546	2 B75573	conserved hypothet
33	101.5	2.5	848	2 C65083	hypothetical prote
34	101	2.5	626	2 T35669	hypothetical ATP/G
35	100.5	2.5	654	2 S76870	hypothetical prote
36	100.5	2.5	1257	1 A41060	neural cell adhesi
37	100.5	2.5	1544	2 E59431	phosphoinositide-b
38	100	2.5	919	2 S42842	T16612.2 protein -
39	99.5	2.5	579	2 E83144	hypothetical prote
40	99.5	2.5	2352	2 C83229	probable non-ribos
41	99	2.5	602	2 T35782	probable secreted
42	99	2.5	915	2 T03589	probable aspartate
43	99	2.5	2055	2 T00093	hypothetical prote
44	98.5	2.5	881	2 S25445	nitrate reductase
45	97.5	2.4	538	2 S76085	hypothetical prote

ALIGNMENTS

RESULT 1

JC5234
amine oxidase (copper-containing) (EC 1.4.3.6) AOC2 Precursor - human

C:Species: Homo sapiens (man)

C>Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 18-Feb-2000

C:Accession: JC5234

R:Zhang, X.; McIntire, W.S.

Gene 179, 279-286, 1996

A:Title: Cloning and sequencing of a copper-containing, topa quinone-containing mon

A:Reference number: JC5234; MUID:97128319; PMID:8572912

A:Contents: Placenta

A:Accession: JC5234

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1763 <ZNA>

A:Cross-references: GB:039447; NID:g1399031; PIDN:AC50919.1; PID:g1399032

C:Comment: This enzyme catalyzes the oxidation of primary amines to the correspondi

Cu(II) and 1 molecule of covalently-bound topa quinone. It also can scavenge circ

C:Genetics:

A:Gene: GDB:AOC2; DAO2

A:Cross-references: GDB:4562632

A:Map position: 17q21-17q21

C:Superfamily: amiloride-binding protein

C:Keywords: copper; metalloprotein; oxidoreductase; quinoprotein; topaquinone

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-763/Product: amine oxidase (copper-containing) #status predicted <MAT>

F:444,520,522/Binding site: copper (His) #status predicted

F:471/Modified site: topaquinone (Tyr) #status predicted

Query Match	100.0%	Score 4010;	DB 2;	Length 763;
Best Local Similarity	100.0%	Pred. No. 1,5e-111;		
Matches 744;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	LVCVLLVGRGGDGEPSQLPHCPSPVSPSACQWTHPGSQSLFADLSREELTVMFRLQRL	60	
DB	20	LVCVLLVGRGGDGEPSQLPHCPSPVSPSACQWTHPGSQSLFADLSREELTVMFRLQRL	79	
QY	61	GPGLVDAQAQAPSDNCVSVSELOLPKRAALAHNDRGSPPARFALIVFGROPOPVVS	120	
DB	80	GPGLVDAQAQAPSDNCVSVSELOLPKRAALAHNDRGSPPARFALIVFGROPOPVVS	139	
QY	121	ELVVGPPLPHPSYMDVVERHGGLPYHRRVLEFOEYLDIDOMTFNRELPOASGLLHHC	180	
DB	140	ELVVGPPLPHPSYMDVVERHGGLPYHRRVLEFOEYLDIDOMTFNRELPOASGLLHHC	199	
QY	181	FYKRGRLVMTTAPRGLSGDRAITWEGLYNNSGAGFLLHNGGLLVNHNKALDPRRW	240	
DB	200	FYKRGRLVMTTAPRGLSGDRAITWEGLYNNSGAGFLLHNGGLLVNHNKALDPRRW	259	
QY	241	TIOKVFYOGGRYDLSAOLEAOFEGAGLVNVLIPNGTGSMSLSPVPPGAPPLQGFYQ	300	
DB	260	TIOKVFYOGGRYDLSAOLEAOFEGAGLVNVLIPNGTGSMSLSPVPPGAPPLQGFYQ	319	

Oy	301	GRFESVQGSVVASLSLMTFSGISGLAFSGPRIFDVROGSRLYVEISLQELALATYIGGNSPAA	360
Dd	320	GRFESVQGSVVASLSLMTFSGISGLAFSGPRIFDVROGSRLYVEISLQELALATYIGGNSPAA	379
Oy	361	MTRTRYDGGGGMKYYTTPLTRGVDCPYLATYVDHMFLLSESQAPKTRDAFCVEQNGCLP	420
Dd	380	MTRTRYDGGGGMKYYTTPLTRGVDCPYLATYVDHMFLLSESQAPKTRDAFCVEQNGCLP	439
Oy	421	LRHHSDLYSHYGGGLAETLVLVRSKSTLNDYVMDTFVHBSGAIEIRFATGYTSSAF	480
Dd	440	LRHHSDLYSHYGGGLAETLVLVRSKSTLNDYVMDTFVHBSGAIEIRFATGYTSSAF	499
Oy	481	LGATGATKYGQVSEHNTLGTVHTHSAHFKVDVDVAGLEKMMVWAEQVFWMAVPMSEHOL	540
Dd	500	LGATGATKYGQVSEHNTLGTVHTHSAHFKVDVDVAGLEKMMVWAEQVFWMAVPMSEHOL	559
Oy	541	ORLOVTRKLLLEMEEOAFLVGSATPRYLYLASNHSNKGHPRGYRIQMLSFAGEPLPONS	600
Dd	560	ORLOVTRKLLLEMEEOAFLVGSATPRYLYLASNHSNKGHPRGYRIQMLSFAGEPLPONS	619
Oy	601	SNMRGFSWMEYUOLAVNORKKEEPSSEVFNONDPAPVPVDSDFINNETIAGDVLAAWYT	660
Dd	620	SNMRGFSWMEYUOLAVNORKKEEPSSEVFNONDPAPVPVDSDFINNETIAGDVLAAWYT	679
Oy	661	AGFLTHIPHAEDIPNTVTVGVNGVGFELRPYNEFDEDPSEYSADSIYFRGODAGACEVNPL	720
Dd	680	AGFLTHIPHAEDIPNTVTVGVNGVGFELRPYNEFDEDPSEYSADSIYFRGODAGACEVNPL	739
Oy	721	ACLPQAAAACAPDLPAPFSHGGSFN 744	
Dd	740	ACLPQAAAACAPDLPAPFSHGGSFN 763	

RESULT 2

amine oxidase: copper-containing (EC 1.4.3.6), serum, precursor - bovine.
C.Species: Bos primigenius taurus (cattle)
C.Date: 09-Sep-1994 #sequence, revision 06-Feb-1995 #text_change 18-Feb-2000
C.Accession: A54411, B38081; A48242; S65408
R.Mu, D.: Medzitradszky, K.F.; Adams, G.W.; Mayer, P.; Hines, W.M.; Burlingame, A.L.; Smith, J.P.
J. Biol. Chem. 269, 9926-9932, 1994
A.Title: Primary structures for a mammalian cellular and serum copper amine oxidase.
A.Reference number: A54411; MUID:94193686; PMID:8144587
A.Accession: A54411
A.Molecule type: mRNA
A.Residues: 1-762 <MUA>
A.Cross-references: GB:S659583; NID:9546215; PIDN:AA830397.1; PID:9546216
R.Mu, D.; Jones, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Kilman, J.P.
J. Biol. Chem. 267, 7979-7982, 1992
A.Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine oxidase.
A.Reference number: A38081; MUID:92235001; PMID:1569055
A.Accession: B38081
A.Molecule type: protein
A.Residues: 463-465, 'D', 467-473, 'X', 475-485 <MU2>
R.Jones, S.M.; Mu, D.; Wenner, D.; Smith, A.J.; Kaur, S.; Maltby, D.; Burlingame, A.L.; Smith, J.P.
Science 248, 981-987, 1990
A.Title: A new redox cofactor in eukaryotic enzymes: 6-hydroxyopa at the active site of copper amine oxidase.
A.Reference number: A48242; MUID:90260648; PMID:2111581
A.Accession: A48242
A.Status: preliminary
A.Molecule type: protein
A.Residues: 468-472 <JAN>
Eur. J. Biochem. 237, 93-99, 1996
R. Blase, D.; Agostinelli, E.; de Matteis, G.; Mondovl, B.; Morpurgo, L.
A.Title: Half-of-the-sites reactivity of bovine serum amine oxidase. Reactivity and chemical mechanism.
A.Reference number: S65408; MUID:96203913; PMID:8620899
A.Accession: S65408
A.Molecule type: protein
A.Residues: 463-469, 'X', 471-487 <DEB>
C:Superfamily: amiloride-binding protein
C:Keywords: copper; glycoprotein; oxidoreductase; quinoprotein; topaquinone
E1-16/Domain: signal sequence #status predicted <STO>

F:17-762/Product: amine oxidase (copper-containing) #status predicted <MAT>
E:136,221,665/Binding site: carboxylate (Asn) (covalent) #status predicted
F:470/Modified site: topaquinone (Tyr) #status experimental

Query Match	83.1%	Score 3331	DB 2	Length 762
Best Local Similarity	81.4%	Pred. No. 2.2e-257		
Matches 609; Conservative	59;	Mismatches 70;	Indels 10;	Gaps 2;

QY	4	VLINGR--GGGGG-----PQDLHCVSPSSAPQMHHPGQSOAFALISHEELTAV	5
Dd	12	LLVMKREGGAGVEBGVKQKCHPSLPRCHSRSPSDQPMHHPDOSQLFADLSHEELTYW	71
QY	54	RELQRLQRLPGVLDAQAARPSDNCVSEYELQ.LPRKAAALAHLDNRSPREAREALATVEFGR	113
Dd	72	SFLQQLQGLGRLDYDAQAARPSDNCVSEYELQ.LPRKAAALAHLDNRSPREAREALATVEFGG	121
QY	114	QOPQVNSLTVYGRPLPHNSYAMDVTVERRGQPLPHNRRPVLFQETLIDQMIFFNRELPOAS	173
Dd	132	QOPQVNSLTVYGRPLPHNSYAMDVTVERRGQPLPHNRRPVLLRETLIDQMIFFNRELPOAA	191
QY	174	GLLHHCCKYKRGRLVMTAPRGLQSGDAPTEFGLYUNISGAGFLLHVGLELLVNHK	233
Dd	192	GLLHHCCKYKRGRLVMTAPRGLQSGDAPTEFGLYUNISGAGFLLHVGLELLVNDHK	251
QY	234	ALDPARMTIQKVFYQGRYYDSLQLEAOFENAGLVNVLIPDNGTGSMSLSPVPGPAP	293
Dd	252	ALDPARMTIQKVFYQGRYYDSLQLEAOFENAGLVNVLIPDNGTGSMSLSPVPGPPTP	311
QY	294	PLQFTRPOGRPSVSGSRVASSLMTFSPGLAFSGPRIFDVAFQGBRLVETLSLOEALATY	353
Dd	312	PLQFTRPOGRPSVSGSRVASSLMTFSPGLAFSGPRIFDVAFQGBRLVETLSLOEAGAVY	371
QY	354	GGNSPAAHTTRTVLDGCFGMKGYTTPRLTGVCNCPRLATYVDMHFLLESQAPRTIDACVF	413
Dd	372	GGNSPAAHTTRTVLDGCFGMKGYTTPRLTGVCNCPRLATYVDMHFLVESQTRKTLIDACVF	431
QY	414	EONQGLPLRRHSDLSHYHFGGLAETVLVYKSMSTLLNYDMVDVHPHSGAIEIRFAT	473
Dd	432	EONQGLPLRRHSDLSHYHFGGAQTVLVERVSSTMLNYDMVDVHPYRNGAIEVUKHAT	491
QY	474	GYISSAFLEFGATGKGNVOYSEHTLGYVTHSHANKVDLDVAGLENWYMAEDNVPVPAV	533
Dd	492	GYISSAFLEFGAARRGNGOVGCHTGLGPRVTHSHANKVDLDVAGLENWYMAEDMAFVPAIP	551
QY	534	WSEHQDRLQVTRKRLMEQOAFVLSA.PRLYLASNHSNKGHPRGRIOMLSFAG	593
Dd	552	WSEHQDRLQVTRKRLMEQOAFVLSA.PRLYLASNHSNKGHPRGRIOTVSEAG	611
QY	594	EPLPQNSSMARQFWMERYQLAVTORKEBEPSSSVFNQNDMPATVYFSDFINNETTACK	653
Dd	612	GMPQNSSMARQFWMERYQLAVTORKEBEPSSSVFNQNDMPATVYFSDFINNETTACK	671
QY	654	DLVAVYTAGFLHIIPHAEDIPNTVTVGNVGFFLRRPYNFEDDEPSFYASDLSYFERGDQAG	713
Dd	672	DLVAVYTAGFLHIIPHAEDIPNTVTVGNVGFFLRRPYNFDDQPSMDASDLSYFERGDQAG	731
QY	714	ACEVNPPLACLPOAAACADPLAFSHGGF	741
Dd	732	SCEINPLACLPOAAACADPLAFVSHGGF	759

RESULT 3

aml1ne oxidase(copper-containing) (EC 1.4.3.6), kidney, precursor - rat
N.Alternate names: aml1oride-binding protein, long form
N.Contains: aml1oride-binding protein, short form
C.Species: Rattus norvegicus (Norway rat)
C.Date: 13-Jan-1995 %sequence_revision 13-Jan-1995 %text_change 18-Feb-2000
C.Accession: S36847; S36848; I51904; S34656; S34657
R.Lingueglia, E.; Renard, S.; Voilley, N.; Waldmann, R.; Chassande, O.; Lazdunski, M.
Eur. J. Biochem. 216, 679-687, 1993
A.Title: Molecular cloning and functional expression of different molecular forms of
A.Reference number: S36847; MUID:93387921; PMID:83375402

Query Match 34.98; Score 1398.5; DB 2; Length 746;
Best Local Similarity 40.38; Pred. No. 3.3e-103;
Matches 294; Conservative 127; Mismatches 266; Indels 43; Gaps 13;

[illegible]

A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1,247, 'E', 249-257, 'V', 260-276, 'I', 277-287, 289, 'I', 291-455, 'P', 457-658, 'D', 6
 A:Cross-references: GB:23670; NID:9474301; PIDN:BA04900.1; PID:9809499
 A:Experimental source: strain K-12; substrain W3110
 R:Steinbach, V.; Benen, J.A.E.; Bader, R.; Postma, P.W.; de Vries, S.; Duine, J.A.
 Eur. J. Biochem. 237, 584-591, 1996
 A:Title: Cloning of the macA gene that encodes aromatic amine oxidase of Escherichia coli
 A:Reference number: S65442; MUID:96235221; PMID:8647101
 A:Accession: S65442
 A:Molecule type: protein
 A:Residues: 41-48 <STE>
 A:Experimental source: strain W3350
 A:Gene: tynA
 C:Complex: homodimer
 C:Function:
 A:Description: catalyzes oxidation of phenylethylamine and water to phenylacetaldehyde,
 A:Pathway: amino acid metabolism; phenylethylamine catabolism
 A:Note: 2,4,5-trihydroxyphenylalanine quinone (topaquinone) cofactor; copper cofactor; h
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: copper; oxidoreductase; quinoprotein; topaquinone
 F:1-33/Domain: signal sequence #status predicted <SIG>
 F:34-757/Product: amine oxidase (copper-containing) #status predicted <MT>

Query Match 9.9%; Score 397.5; DB 2; Length 757;
 Best Local Similarity 27.5%; Pred. No. 2.5e-23;
 Matches 142; Conservative 74; Mismatches 211; Indels 89; Gaps 20;

QY 246 FVQGRYDLSLQLEAFEGVNVVLIIPNGTGGSS-----LKS 285
 DB 259 YEDGK--DELKQ-----DKLLKLVISYLDVGDGNTYANHPLENVAVYDLEQKIVIEEG 311
 QY 286 PVPKPG-----AP-----PLOGY--POGPRFSVQSRVASSLMTSFGIGAFSGPR 329
 DB 312 PVPVPMATAPFGDGRVAPVAPVPMQIIEPEGKNTITGDMIMRMDFLMSNRVGP 371
 QY 330 IFDVRFQG-----ERLYEISLQALAIYGGNSPAAATRTRYVGG--FGMKYTTPLTRGV 384
 DB 372 ISTVYNDNGTKRKRYEGSLGMIIVYGDPIGWTFKALDSDGDMGTLSPIARGK 431
 QY 385 CPPLATVYVDMHLELSQAPRTIRDAFCVFQNGOLPLRRHSDLYSHFG-----LAET 439
 DB 432 APSNAVLNLTADYGVMEIRPRLAVFE-----KAGREYKQENQGPVNSERR 483
 QY 440 VLVVRSSTLNDYVDYFHFSGAIEIRFATGY-----ISSAFLEGATG---YGN 490
 DB 484 ELVYRMTSTGVNDYIFEDMIFHENGSTIGDAGATGEAVGVAKTWHDEAKDTRGT 543
 QY 491 QVSEHTLGVTHSAHFKVDLDVAGLENNVMAEDMVFVPAVVPSPRPHQRLQDLYTKL 550
 DB 544 LIHNIYGTGTHOIHYNFRDLDDGNNNSLVAMPVYKPRNT---AGSPRSTGMVQNYNI 600
 QY 551 EMEQAAFLVGSATPRYLYLASNHSNKGPRGRIOMLSFAG--EPPLQNSAA----- 603
 DB 601 GNEDDAQKDFDGTIR--LSPNPKENKGNVSI--QIIIPAGSTHYAKAQAQAPRDEMI 657
 QY 604 -RGFSWERYOLAVYQKKEEPPSSSVF--NONDPWAPTVDSDFTINNETIACKDLVAVTA 661
 DB 658 YHRSEFDKDLWYTRYHGRFEGKYPNRSTHDTGLQYSK--DNESLDNTDAVVMWTT 715
 QY 662 GFLLIPIHAEDIPNTVTVGNGVGFLLRYNPFDEPS 697
 DB 716 GTTHVARAEWPIIMPV--EWVHTLLKPMNFEDEPPT 749

RESULT 7
 A56102
 amine oxidase (copper-containing) (EC 1.4.3.6) - Arthrobacter globiformis
 C:Species: Arthrobacter globiformis
 C:date: 13-Apr-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 C:Accession: A56102
 R:Choi, Y.H.; Matsuzaki, R.; Fukui, T.; Shimizu, E.; Yorifuji, T.; Sato, H.; Ozaki, Y.;

J. Biol. Chem. 270, 4712-4720, 1995
 A:Title: Copper/topa quinone-containing histamine oxidase from Arthrobacter globifo
 A:Reference number: A56102; MUID:95181469; PMID:7876243
 A:Accession: A56102
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-684 <CHD>
 A:Cross-references: GB:D38508; NID:9994746; PIDN:BA07517.1; PID:9994747
 A:Note: authors translated the codon TIG for residue 239 as phe
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: oxidoreductase

Query Match 9.4%; Score 377; DB 2; Length 684;
 Best Local Similarity 24.1%; Pred. No. 9.1e-22;
 Matches 182; Conservative 107; Mismatches 343; Indels 124; Gaps 28;

QY 20 PRCPSVPSAQWMT--HPCQSOLFADLSREELAVVRFVLTORLGLVDAQAARPSNCV 77
 DB 7 PESTPLVQDPVPVATLVHAAQHPLELSAEIHEARRILAE--AGLVG----- 52
 QY 78 FSEVQLPPKAALAHLDNGSP-----PAREALAYFPGRPQ-----PNVSE 121
 DB 53 -----ESTRFAYLGLIEPPKTRTGQDVYGAARLVAMLPAAQSSLDVRLSLATG 103
 QY 122 LVVGPLPHPSYMRDVVERHGGPLPYHRRPVLEFOEYLDIDQMFNRELPOASGLHNHCF 181
 DB 104 LVVD-----RRELNPEDAG-----QLPVLEEGLIEDL--SEDPQMNAL----- 143
 QY 182 YKHGRNLYMTAP-----RGLOSGDRATWGLYVNSGAFGLHNGVLELVNHNKA 234
 DB 144 -TARGTLPAQVAVRPAVSAGVEFYGNEGKRL-----LRGSGFRQDHP-----ADHPW 189
 QY 235 LDPARNTIQVYFQGRYDLSLQLEAFEGVNVVLIIP--NGTGSWSLSKSPVQGPAP 293
 DB 190 AHPIDLVAVDVENRRVNL-----IDGSPVAV--PEVNGYTDPRALRGELRIDL 240
 QY 294 PLOGYPOGPRFSVQSRVASSLMTSFGIGAFSGPRIFDVR--FOGER--LYEISLQEA 349
 DB 241 IETMOPBGFPTLEGNLSMAGNDLRVGPAREGLVHQHNSHKGRRRPIYHRAISSEM 300
 QY 350 LAIYGGNSPAAATRTRYDGG--FGMKYTTPLTRGVDCPYLATIYDMHFLLESAQPTIR 408
 DB 301 VVPYGDSPYRSMQNYFDSEYLVGRDANSRLRCDCIGITVYNSPVVADDFGNPRTEEN 360
 QY 409 AFCVFQNGOLPLRRHSDLYSHFGSLAEVLYVVRMSSTLNDYVDYFHFSGAIEI 468
 DB 361 GICIHEDAGILMK--RTDEWAGSDEVRRNRRLVVSFFTYGVNDYGFYTWLYLDGTIEF 418
 QY 469 RFVATGYISSAFULGATGKGNQVSEHTLGTVTHSAHFKYVLDVAGLENNVMAEDMVF 528
 DB 419 EAKTGVIFTRALPDKRYATVASELAPGAGAYHOLHSARIDMMIDGQANVEELDLYRL 478
 QY 529 PMAVVPSREHOLQVYTRKLEMEQAAFLVGSATPRYLYLASNHS--NKGHGRGRIQ 587
 DB 479 PKG--PGNPHG--NAFTQKRTLLARESEAVRDADAKGVWISIPDSLNHIGHPVGYLY 535
 QY 588 MLSFAGER---LPONSSMARGFSWERYOLAVYQKKEEPPSSSVFNONDPWAPTVDSEDF 644
 DB 536 P--EGNPTLAMADSSIASRAAFARHNLWTRNAEEELVYAGDFVNOHPCGAVLP--AYV 591
 QY 645 INNETIAGKDLVAVWTAGFLIPIHAEDIPNTVTVGNGVGFLLRYNPFDEPSFYSADSI 704
 DB 592 AQDDIDIGODLVVHNSFGLTHFRPREDMP--IMPVDTGFLTKPGRGFDEPNTLVVPSA 649
 QY 705 YFRGDQAGACEVNPACLPOAAACAPDLPAFSHG 740
 DB 650 -----AGHCGTG-----SERENHAPGTAIVGHSG 673

RESULT 8
 G90330
 amine oxidase (copper-containing) (tynA) [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: G90330
R:Shen, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Ameyez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arlett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
Submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: G90330
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1,660 <KUR>
A:Cross-references: GB:AE006641; NID:g13814942; PIDN:AAK41902.1; GSPDB:GN00155
C:Genetics:
A:Gene: tynA
C:Superfamily: amine oxidase (copper-containing)

Query Match 9.2%; Score 368.5; DB 2; Length 660;
Best Local Similarity 24.4%; Pred. No. 4,1e-21;
Matches 175; Conservative 103; Mismatches 303; Indels 137; Gaps 28;

QY 44 LSRELTAVMRFLLQRLGPGLVDAQAARPSDNCVFSVELQPPKAAALAHLDRCSPPAR 103
DB 17 LNEEIKRSVEVLKROLN---LDAKVKV---FFSVELKEPKKOEYLEWNNKNIKIER 68
QY 104 EALIVFEGROPQPNVSELVGPRLPHPSYMDYVERHGSGFLPHRRPVLFQETLIDQM 163
DB 69 ESL-VKYYNFODR-KVYEAL---SLDNNVYKEIRSIDAA--PPTLDEFCECECA 118
QY 164 IFNRELPOASGLLHCCFYKRGHRLVMTTAPRGLOSGDRATFGLYINISGAFLLH 223
DB 119 VRNKRQGEA-----LTKGILINDLN---LWDCWABP---H 151
QY 224 VGELELVNKAALDPAKTIQKVFYQGRYDSLAOLEAFAGLVNVLIPDNGTGSWSL 283
DB 152 VDEELRGRVAIG-YMW-VKKDIEDNGYGRVHGIMPVLDKMEVIRIDHGT----- 203
QY 284 KSPVPPPA-----PPIQF-YRQGRFSVQGSRAVSSLTWTSFGLGAFSG 327
DB 204 -SPRLDPAVYTPKCLKIGDGLKPIEIKOPLSSSIKINWEISYRMLRLIYTPREG 262
QY 328 PRLEDFVRQGER-----LVYEISLQELALAYGNSPAAMTTRYVDG-FGCKYTPPLTR 381
DB 263 LVYIDVYIDENNERRMILYRASVVDLWVPGDPSPHNKKMVLADAGDYGIGNTIVPLSH 322
QY 382 G-----VDC-PRYATVVDHFLLESQAPKTRIDAFCEVQONGLPLRRHSDLYSHYFG 434
DB 323 GNYDLVNCDFGEVITYHLDTVRVSSNGTPIKIKKACVHEEDFGVLMR--HTDLRS--- 376
QY 435 GLAET-----LVVYRSMSTLNYDYVMDTVFHPGATIEIRYATGYISSAFL--GATGKY 488
DB 377 GKSEVRNRRLVYSFMATLANYDGFYFWYODGSIEFLVKLGTIINDDSISEKDPTRKY 436
QY 489 GNOVSEHTLCTVHTSHAFKVDLDVAGLENNVMAEDMVFVMAVPSPEHOLQRLQYTRK 548
DB 437 GTRVTPEVYAPRIHHEFNIRLINVGLRNRIT-----EVLNREBPITEK 481
QY 549 -----LLEMEQAAFLVGSATPRYLILAS--NHSNKGCHPRGYRIOMLSFAGEP 595
DB 482 NPVGNAFPAEENLEENADARRHVNPDQTRGWKIVNIQKNKYLGLPYAYRLVPGHNVLP 541
QY 596 LPONSSVARGFSMERVQLAVTQKKEEPPSSSVFNQNDMPAAPTDFSD-----FINNETI 650
DB 542 LPDDSYRRGAGATINHLMTVYNEERVYASGD-----PYLRADGGLPKYTLKKRSI 594
QY 651 AGDLVAMVYAGFLIHPAEDIPTVTVGNGVGFLLRPYNFEDDPFSYASDSIYFRG 708
DB 595 VDEDLVIMVTLGVENHYRIEDMP--VMPVEMAGFRLIPDFFPKNPIIYLPROLRING 650

RESULT 9
A48646

amine oxidase (copper-containing) (EC 1.4.3.6) - Arthrobacter sp. (strain P1)
C:Species: Arthrobacter sp.
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999
C:Accession: A48646
R:Zhang, X.; Fuller, J.H.; McIntire, W.S.
J. Bacteriol. 175, 5617-5627, 1993
A:Title: Cloning, sequencing, expression, and regulation of the structural gene for t
hylotriph.
A:Reference number: A48646; MUID:93374858; PMID:8366046
A:Accession: A48646
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1,648 <ZHA>
A:Cross-references: GB:L12990; NID:g289152; PIDN:AAA22074.1; PID:g289153
C:Superfamily: amine oxidase (copper-containing)
C:Keywords: oxidoreductase

Query Match 9.1%; Score 363; DB 2; Length 648;
Best Local Similarity 24.0%; Pred. No. 1,1e-20;
Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

QY 36 GOSQLFADLSREELTAVMRFLLQRLGPGLVDAQAARPSDNCVFSVELQPPKAAALAHLD 95
DB 12 GVSHPDLPSLVELARAVAILKE--GPAAESFR-----FISVLEKPSKDL----- 57
QY 96 RGSPPARREALAIVFEGROPQPNVSELVGPRLPHPSYMDYVERHGSGFLPHRRPVLFQ 155
DB 58 RAGVAVAREADAVLV--DRAQSRSEAVVD-----LEAGTVDSWKLLAENIQDPFLD 108
QY 156 EYLDIDQMIFNRELPOASGLLHCCFYKRGHRLVMTTAPRGLOSGDRATFGLYINISGAF 212
DB 109 EFACEDAC--RKDE-----VIALAKRGILNLDVCEBPMSVGYF 148
QY 213 NISGAGFLLHVGLELVNKAALD-PAKTIQK--VFYQGRYDSLAOLEAFAGLVNVLIPDNG 269
DB 149 GEDNEGRRL--MRLVYPRDADSDSPYAPHTENIVFY-----DLNAG--KV 191
QY 270 VLPDNGTGSWSLSKSPVPPGAPPLQFY-----PQGRFSVQGSRAVSS 314
DB 192 VRLDDQA-----IPVSARGNVLPKYVGEARTDLKPLNITQPEGASFVTGNHTVMA 244
QY 315 LMTFSFGIAGFSRIFIDVRQGE-----RVYELISLQELALAYGNSPAA--AMTRYVDG 369
DB 245 DMSFRVGTTPREGVLVHQLKFKDGVDRPVINRASLSMNVVYDGTAVQAKKAAPDSGE 304
QY 370 FGMRKYTTPLTRGVDCPLATVYVDMHFLLESQAPKTRIDAFCEVQONGLPLRRHSDLY 429
DB 305 YNIGMANSLTLGDCDCEIKYFPGHSDVSHGNWTIENALCMHEEDSI-----LW 356
QY 430 SH--YFGGLAET-----LVVYRSMSTLNYDYVMDTVFHPGATIEIRYATGYISSA-FLF 482
DB 357 KHDFRGTATETRSRKLIVISFIATVANYEYAFWMHFLDGSIEFLVKATGILSTACQLP 416
QY 483 GATGKYGNOVSEHTL-GVHTSHAFKVDLDVAGLENNVMAEDMVFVMAVPSPEHOL- 540
DB 417 GEKPYQOSLNDGLVAPRIHHEFNIRLINVGLRNRIT-----EVLNREBPITEK 481
QY 541 -QRLQVTRKLLLEMEQAAFLVGSATPRYLILASNHS--NKGCHPRGYRIOMLSFAGEPLPQ 598
DB 469 GTAFMAVDRLLETQKAIKRTNEAKHFRFKIANHESNLVNEPARYL-----IPT 519
QY 599 NSSVARGFSMERVQLAVTQKKEEPPSSSVFNQNDMPAAPTVD-----FSDFINNET--- 649
DB 520 NG-----IOLAA--RDDAVYSKRAQFARNMLVTAADYTRERPAAGEYPMQATGAD 567
QY 650 -----TAGDLVAMVYAGFLIHPAEDIPTVTVGNGVGFLLRPYNFEDDPFS 697
DB 568 DGLHITQKRNINVDVTLVVMYTFGHHVYVLEDMP--VMPRONIGMLPEHGFENQNP 625
QY 698 FYSADSIYFRGDODAGAC 715
DB 626 LNLPTSTTGTGTGADTC 643

RESULT 10

copper amine oxidase-like protein - Arabidopsis thaliana
T48139
N:Alternate names: protein T4C9.130
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
C:Accession: T48139
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Men
submitted to the Protein Sequence Database, June 1999
A:Reference number: 224485
A:Accession: T48139
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-756 <BEV>
A:Cross-references: EMBL:AL080318
A:Experimental source: cultivar Columbia; BAC clone T4C9
C:Genetics:
A:Map position: 4
A:introns: 460/3; 498/3; 686/2
A:Note: T4C9.130
C:Superfamily: amine oxidase (copper-containing)

Query Match 8.93; Score 357.5; DB 2; Length 756;
Best Local Similarity 22.28; Pred. No. 3.0e-20;
Matches 161; Conservative 100; Mismatches 280; Indels 191; Gaps 29;

QY 75 NCVSVELQLPKAAALAHLDGSPPPAREALAI-----VFGRQDPQPNVSE 121
D 126 HALHTVLEBEK NLYRHKGNPLPRKASVARIAGDHTLVLDISTGRVDSENSPV 184
QY 122 LVVG-PLPHBSYMDVVERHGGPLPYHRRPVLFQELVDIDQMFNRELPAASGLHHCC 180
D 185 RVSGYPMWTIEEMNDIV-----VPSNAD-----FNRTIIS----- 216
QY 181 FYKRGRLVMTAPRGLSGDRATWFG-----LYYNISAGFELHHV-GLE 227
D 217 ---RGVNLIDVICFP--ISCG---WFGNKEENARKVKSOCFMTQGTPTNYMRIDELT 266
QY 228 LVNKKALDPAKWTIQKVFYQGRYDLSLAOLEAFGLVNVVLLIPDNGTSGSWSLSPV 287
D 267 ILID--LDTKQ-----VIEITDGRA-----IPI 288
QY 288 PPG-----PAPLQF-YPOGPRFSVQSGRYAS-SLTFPSGLCAFSGP 328
D 289 PGSTNTDYRFQKLAITTDKTRPLNPISIEQPRGPFVIEDNHLVKKANNEFLKPPRAGV 348
QY 329 RLFQVR-----FQGERLYVEISLOALAIYGNSPAMTTRYYVG-GFMGKTYTPTLR 381
D 349 VISRRVADPDTHERDVTYKGFSELFVPIYMPDPAWFKTYMDAGEYGFQLOAMPVLP 408
QY 382 GVDCPYLATYVDMHFLLESQAQKTRDAFCVFEQN-----QGLPRRHNSDL 428
D 409 LNDCCRNAAKYMDGVFAADGTPFVRENMCIESYAGDIGRHSSEPIGIDIREVR--- 465
QY 429 YSHYFGGLAETLVLVRSMTLINDYVMDTVFHPGALDIRYAAIGYISALFQATGKY 488
D 466 -----PKYTLVVRMAASVGNNDYIIDYEFQTDGLIKAKGLSIL---WVKGTYYON 514
QY 489 GNOV-----SEHTLGTVHTSAHFKVDVAGLEVMVAEDMVEFPMAY 532
D 515 KQVQEKDKDGNNEELHGLTISENVIGVINDHTVTFYLDLDVGPDPN-----SFVKYNLKR 569
QY 533 PMSPEHQLQR---LQVTRKLEMEQAAFLVGSATP-RYLYLASNHNKMGHPRGRIOM 588
D 570 QETEGESPRKSYLAKVARIATKENDQIKLSLYDSEEHVINSKGTITVGNPTGYKVV 629
QY 589 LSFAGEPLPONSSMARGFSEWERYOLAVTORKEEPESSSVFNQNDPMAPTVD-FSPFINN 647
D 630 RTTASLSDHDPQKRGAFITNQIIVWTFYNNKSEQWAGLFTYQSHGDDTLAWMSD--RD 687
QY 648 ETIAGCDLVAVWTAGFLIHPHADDIPNTYVYGCVGFILRPVNFPEDESPFSADSIYR 707

Db 688 RDIENKDIWVMTLGFHHIPQEDPEIMPVSS--SFDLKPVNFFERNPILSAAPNF--- 742

QY 708 GQDDAGACEYNPLA 721

Db 743 -EHDLPVCGVQSVS 755

RESULT 11

probable amine oxidase - Arabidopsis thaliana
G71412
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C:Accession: G71412
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.;
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.;
avanagh, T.; Hempel, S.; Kotter, P.; Entlian, K.D.; Rieger, M.; Schaeffer, M.; Funk,
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgido
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reichman, S
C.; Chalmatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsi
A:Reference number: A71400; M01D:98121113; PMID:9461215
A:Accession: G71412
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-650 <BEV>
A:Cross-references: GB:297337; NID:q2244829; PID:q2244851
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: amine oxidase (copper-containing)

Query Match 8.83; Score 353.5; DB 2; Length 650;
Best Local Similarity 23.98; Pred. No. 6.3e-20;
Matches 161; Conservative 96; Mismatches 278; Indels 139; Gaps 29;

QY 80 VELQPPKAAALAHLD---DRGSPPPAREALAIFFGRQDPQPNVSELVGLPHPSYRD 135
D 56 LDLEPPKNSHLQMLSPKSPKPPRRRSRYVYVAGQ---TYELI-----D 101
QY 136 TVTER-----HGGP---LPYHRRPVLFQELVDIDQMFNREL--PQASGLHHCC 180
D 102 LTTSKIASRIYTGHPFSFTFIELEFKASKLPJLTPPKKSLDLSLISEVSCIPFTVG 161
QY 181 FYKRGRLVMTAPRGLSGDRATWFGLYYNISAGFELHHV-GLELVNKKALDPA 239
D 162 WYGE-----TTTRELKA-----SCFTRDGSVNVFTPIGITYTID--VD--- 200
QY 240 WTQKVFYQGRYDLSLAOLEAFGLVNVVLLIPDNGTSGSWSLSPVPPGPAPLQFY- 298
D 201 -SMQYIKISDRKRP-----IPK-BENDRRTK-----RFPFFC 234
QY 299 -POGPRFSVQSGRYASSLTFSEFGLCAFSGPPIFDV-----RQGERLYVEISLOEA 349
D 225 NVSDTGFKILGRVKKAMKFFHVGFTARAGVITISASVLDPRTKRR--RVNRRHVSET 292
QY 350 LAIYGNSPAAATTYYVD--CGFGMGKTYTPLRQVDCPYLATYVDMHFLLESQAQKTRID 408
D 293 FVPIYMDPTYEMWYRTFMDIGFEGFGSAVNLPLDPCPNAAFLDGHVAGPGGTQAKMTN 352
QY 409 AFCVEQNGQLPLRRHNSDL---SHYFGLAETLVLVRSMTLINDYVMDTVFHPGA 465
D 353 VMCVEFEKN-GYASAPRHTIINPGVITISGEAELISLVYVMVATLIGNYDVIWEEFKNGA 411
QY 466 IEI-----RFATGYISSAFLECATKRYGNQVSEHTLGVYHTHSAPFKVDLVAGLE 517
D 412 IRVGYDLTGVLLEVYKATSYNSDQI--TENNYGTGLVAKNTIAVNHHDYLLYYLDLDVNG 469
QY 518 NNVMAEDMVFYVMA-VPNAPPEHQLQVOTRKLEMEQAAFLVGSATPRTYLYLASNHN 576
D 470 NSLVKAKLKTIVRTVEYNNKSSRRKSYWTVYAKLEADGRVRLGSDPVPELLIVNKKKT 529
QY 577 KMGHPRGRIOMLSFAGEPLP---QNSSMARGFSEWERYOLAVTORKEEPESSSVFNQND 633

Db 530 KIGNTVGYRL-----IPEHLQATSLITDDDDPELRAGY-----KRPVWVTVLRNSE 576
 Qy 634 PWAPVDS-----FINNETAGKDLVAVMTAGFLHIAEDIPNTVTVNGCV 682
 Db 577 RWAGGF-YSDNSRGDDGLAVSSNRREIENKDIWVMYNGFHIIYOEDEFVPTLHG-- 633
 Qy 683 GFELRPVNFEDDP 696
 Db 634 GFTLRPSNFNDP 647

RESULT 12

JC2139
 Phenylethylamine oxidase (EC 1.4.3.-) - Arthrobacter globiformis
 C:Species: Arthrobacter globiformis
 C:Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 28-May-1999
 C:Accession: JC2139; PC2070
 R:Tanizawa, K.; Matsuzaki, R.; Shimizu, E.; Yorifuji, T.; Fukui, T.
 Biochem. Biophys. Res. Commun. 199, 1096-1102, 1994
 A:Title: Cloning and sequencing of phenylethylamine oxidase from Arthrobacter globiformis
 A:Reference number: JC2139; MUID:94197690; PMID:8147851
 A:Accession: JC2139
 A:Molecule type: DNA
 A:Residues: 1-638 <TAN>
 A:Cross-references: GB:U03517; NID:9451488; PID:AA18114.1; PID:9451489
 A:Accession: PC2070
 A:Molecule type: Protein
 A:Residues: 3-18;38-52;132-142;185-209;219-235;402-419;474-497;504-537;540-559;611-625
 C:Comment: This enzyme catalyzes the oxidative deamination of various biogenic primary
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: copper binding; oxidoreductase; quinoprotein; topaquinnone
 F:355,431,433/Binding site: copper (His) #status predicted
 F:382/Modified site: topaquinnone (Tyr) #status predicted

Query Match 8.58; Score 340; DB 2; Length 638;
 Best Local Similarity 23.7%; Pred. No. 7.4e-19;
 Matches 157; Conservative 100; Mismatches 274; Indels 132; Gaps 28;

Qy 87 KAALAHLD--RGSPPPAREALAIYF---GROPQVNSLVLPRLHPSMDVTER 140
 Db 38 RAVLVGVDPARGAGSEEDRRFRFVHDVSGARQ-----EVTYSV 79
 Qy 141 HGGPL-----PYHRRPVLFQETLIDIDIMFN-----RELPOAS 173
 Db 80 TNGVYISAVELDTAATGELPVLEEFVEVDLATERMLKALARNLDVSKVRAPLSA 139
 Qy 174 GLHHCCYKRRGRNLYMTAPRGL---QSGDRATWFLGYLINSAGAGFLHHVGL 229
 Db 140 GVFEYA---EERGRIL-----RGLAFVQDPEDPSAW---AHPYDGLVAVYDVYSKEVT 187
 Qy 230 VNHKALDPARTIQKVFQGRYVDSLAQLAQLFAGLVNVVLPDNGSGSLSKSPVP 289
 Db 188 ---KVIDGVFVPA--EHGNTDP-----ELTGPLRT 215
 Qy 290 GRAPDLOFYPGPRFSVO--GSRVASSLMTFSGLCAFSPRIFDVRFQ--GERL---VVEI 344
 Db 216 TQKDISITQPGSPFTVGNHIEKWSLDVGFVGRVVLHNAFQDGRRLRIINRA 275
 Qy 345 SLOGLALAYGNSRAAMTRKYD--GGGGMKITYPLTRGVDC---PLYATVDMHFLLE 399
 Db 276 SIAMVVPYGPSPISRMQNFDTGEYLVQYANSLGCGCLGDTITLSPVISAQF--- 332
 Qy 400 SQAKTIRDACVFEONQGLPLRRHSDLYSHYFGGLAETVLVYRMSSTLINDVVMPTV 459
 Db 333 -GNREIRNGICMHEEDMGLAK--HSDLMGICNTRNRKRVISFTTIGYVDGDFWY 389
 Qy 460 FHPGSAIEIRYATGYI--SSAFLEFAGTKYGNQVSEHTLGTVHSHAFKVDLVAGLEN 518
 Db 390 LYLDGITFEAKATGAVTSAFPEGGSQNI--SQLAPLGAFPHQHIFGARLMDAIDGFTN 448
 Qy 519 WMAEDVVFVMAVPMSEHQLQVTRKLLMEEQAAFLVGSATPPTVYLASHNS--NK 577

Db 449 RVEEDVVRQTMGPENRGNAFSR---KRTVLRSEAVREADARTGRTWITSNPSKNR 505
 Qy 578 WCHPRGIRI---QMSFAGEPLPQNSMARGFSEWRYOLAVTORKEEPPSSSVF--NND 633
 Db 506 INEPVGYKLLAHNQPTLLADP---GSSIARRAAFKTKLMTWRYADDERPYFGDVNHS 562
 Qy 634 PWAPVDSDFINNETINGKDLVAVMTAGFLHIAEDIPNTVTVNGCVGLFRLRYNFD 693
 Db 563 GGAGLPST--IAQDRDIDGQDIIWMTGTLHPVEDMP--IMPVDTGFLRPEGFD 618
 Qy 694 EDP 696
 Db 619 RSP 621

RESULT 13

JC7251
 amine oxidase (copper-containing) (EC 1.4.3.6) - garden pea
 C:Species: Pisum sativum (garden pea)
 C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
 C:Accession: JC7251
 R:Koyanagi, T.; Matsumura, K.; Kuroda, S.; Tanizawa, K.
 Biosci. Biotechnol. Biochem. 64, 717-722, 2000
 A:Title: Molecular cloning and heterologous expression of pea seedling copper amine o
 A:Reference number: JC7251
 A:Accession: JC7251
 A:Molecule type: mRNA
 A:Residues: 1-674 <KOY>
 A:Cross-references: DDBJ:AB026253
 A:Experimental source: seed
 C:Comment: This enzyme, a homodimer, containing one Cu²⁺ ion and one 2,4,5-trihydroxy
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: copper; copper binding; homodimer; oxidoreductase

Query Match 8.18; Score 325.5; DB 2; Length 674;
 Best Local Similarity 24.18; Pred. No. 1.1e-17;
 Matches 146; Conservative 85; Mismatches 233; Indels 141; Gaps 30;

Qy 158 LDIDIMINRELPOASGLHHCCF---YKRGRLVMTAPRGLQSGDRATWGLYNI 214
 Db 128 LSVDEOSLAIEIP---LKYPPFDSYVKRGLNLSEIYC-----SSFTMGWGEKKNV 176
 Qy 215 -----SGAGFLHHV--GLELVNHNKALDPARTIQKVFQGRYVDSLAQLAQL 262
 Db 177 RTVRLDCFMKESYVNIYRPTGITVAD---LD-----LKKIV---EYHR----- 217
 Qy 263 EAGLVNVVLPD--NQTGGSWSLSKSPVPPAP--LQFYPGPRFSYVSGSRVASSLMTFS 319
 Db 218 -----DIAVPTAEKTEYQVSKSP--PRGPKHSLTSHPQPGPQIEGHSVANMKFH 271
 Qy 320 FGLGAFSG-----PRIPV--RFGGERLYEISLOBALAYGNSRAAMTRKYVDG--FGM 372
 Db 272 IGFVRAQIVISLASIYDLEKHSRVLVYKGYISLFPYQDPTPEEFYKFFPDSGEFGF 331
 Qy 373 GKTYPLRAGVDCPLATVYDMHFLLESQAPKTRDACVFEONQGLPLRRHSHDLYSHY 432
 Db 332 GLSTVSLIPNDCRPHAFQIDITVHSANGPILKNAICVEQ-----YGINMHT 383
 Qy 433 FGL-----AETVLVRSMSSTLINDVVMPTVFHPGSAIEIRFYATGYISAFLEG 483
 Db 384 ENGIPNESIEESRFEVNLIVRTIYGVGNVDVIMEFASSINPALASQILE---IKG 440
 Qy 484 ATGK-----YGNVSEHTLGTVHSHAFKVDLVAGLENWMAEDVVFVPM----- 530
 Db 441 TNIRKDEIKEDLGKLVANSISGIYHDFIYVLDPIDIDTHNSFEKTSIKTVRIKGS 500
 Qy 531 ---AVPMSPEHQLQVTRKLLMEEQAAFLVGSAPRYYLYLSNHSNKKGHPGRYRIQ 587
 Db 501 SKRSYWTTEQTAK-----TESDAKTTIGLAPELVVPNPNTKAVGNEVGRFL- 550
 Qy 588 MLSFAGEPL-----PQNSMARGFSEWRYOLAVTORKEEPPSSSVFNQNDPWAP--TV 639

Db 551 IPAIRAPHLFEDDYPOQ-----IRG-AFTNYNWT-----AYNRETKMAGLXY 594
Qy 640 DFS---DFT-----NNETAGKDYAMVATAGLHIAHDIPRTVYGVNGVGFLLRPYNF 691
Db 595 DSRGDDTLAVATKONREIVNKDIYMHVYGIHVPADQDFLIMPL--STSELRPTNF 652
Qy 692 FDEDP 696
Db 653 FERNP 657

RESULT 14
747403
amine oxidase-like protein - Arabidopsis thaliana
N:Alternate names: protein F23N14.50
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C:Accession: T47403
R:Blocker, H.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24465
A:Accession: T47403
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <BIO>
A:Cross-references: EMBL:AL138638
A:Experimental source: cultivar Columbia; BAC clone F23N14
C:Genetics:
A:Map position: 3
A:Introns: 145/3; 303/1; 402/3; 440/3; 591/3; 624/2
A:Note: F23N14.50
C:Superfamily: amine oxidase (copper-containing)

Query Match 8.1%; Score 324; DB 2; Length 687;
Best Local Similarity 22.1%; Pred. No. 1.6e-17;
Matches 166; Conservative 102; Mismatches 298; Indels 192; Gaps 31;

Qy 20 PHCSVSPSAQPTWHPQSOQLFADLSRELTAVNRFLTQRGLVDVAQAARPDNCVFS 79
Db 39 PHP-----LDPLTTP-----EIKRVOTILSGH-DPFGSGS-----TIHA 74
Qy 80 VELDLPKAAALAHLDGSPAREALATVFFGRQPOPNV---SELVVGPLPHPSY---- 132
Db 75 MALDEPKORVIRKKKGRDLPRRAELIANGSGSHVLTDLKGRVYSDLVNTPFGPI 134
Qy 133 --MRDVTVERHGRLPYHRRPVLFEQYLIDIDQMTFNRELPOASGLLHHCFFYKRGRLV 190
Db 135 LTMDDIAV---SQVPY-----KSYEFNRISIEARGIP-FSGLICITPPAGWYGPD-- 180
Qy 191 TMTTAPRGLDGDRAITWFGLYYNSGAGFLHHV-GLELLVNHKALDARMTIQVFG 249
Db 181 -----EGRARYIKQCFSKQDVTNFMPIEGLLTDMKLE----- 218
Qy 250 RYDLSLAQLAQFEAGLVNVLIPDNGGSGMSLSPVPPGAPAPLP----- 297
Db 219 -----IKIYDNG-----PVYPKSGTGTFRYRGFLNETYVYMRV 252
Qy 298 -----YPOGRFRFSVO-GSRVASSLMTFSGLGAFSGRIFDVRFQGERLYEISLOEAL 350
Db 253 NPMSEQDFGSPFVEDGYLVKMANMKFHKPDORAGMI-----SOATVADSKTGEAR 306
Qy 351 AI-YGNSPAM-----TTRYVDG-GFGMKXTPTPLTRGVDCPYLATYVDHFL 397
Db 307 SVMKGFASSELVPPNMDGEGWYSKAYMDAGEFGSPSSMLVPLNDCEPRNAIYTDGFFA 366
Qy 398 LESQAPKTRIDAFVCEONOGPLRLRHSDLSHYFGGL-----AETLVVRSNSTLN 451
Db 367 SPEGIPIIOPNNICFERYAGDTSWRHSIL---LPGVDIESRAKATLVARMASGVN 422
Qy 452 YDYVMDTVFHPSGAIEIRFYATG-----YISSAFLEGATGKYGNOVSEHTLGTWTH 503
Db 423 YDIIDMEFQMDGVIRVTVAASGMILVKGATAYENVEDLGEKEDSDGSLISENVIGVVDH 482

Qy 504 SAHFVLDVAGLENNVMAEDNVFVMAVPWSPEHOLR---LQVTRKLEMEQAAFLV 560
Db 483 FISFHLMDIDGSAN-----NSFYKVLHEKORLPPGSRKRSYLKVKYAKTEKDAQIRM 538
Qy 561 GSATPRYLILAS-NHSGNKGHPRGYRIOMLSFAGEPLPQNSSARAFSEMEROLATYORK 619
Db 539 SLVDPEFHLVDPNRLSLGPNAGYKLVPGGNASILLDDHDPQMRGATFNQIWTNR-- 596
Qy 620 EEEPPSSSVFNONDPWA-----PTVD-FSDFINNETIAGKDYAMVATAGFLHIP 667
Db 597 -----YNRSEQWAGGLIMYQSRGEDTLQWSD--NRSTENNDIYMTLWLGTHHP 645

Qy 668 HAEDIPNTVYGVNGVGFLLRPYNFDEDP-----SFYSAD 702
Db 646 QGEDFPVMTPLAS--SPELKPVNFEEFNPLYGLSPFEKD 683

RESULT 15
C44239
amine oxidase (copper-containing) (EC 1.4.3.6) precursor - garden pea
C:Species: Pisum sativum (garden pea)
C>Date: 31-Dec-1993 #sequence_revision 05-Jan-1996 #text_change 17-Mar-2000
C:Accession: A57327; C44239
R:Tippling, A.J.; McPherson, M.J.
J. Biol. Chem. 270, 16939-16946, 1995
A:Title: Cloning and molecular analysis of the pea seedling copper amine oxidase.
A:Reference number: A57327; MID:95348126; PMID:7622512
A:Accession: A57327
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-674 <TIP>
A:Cross-references: GB:L39931; MID:9685197; PID:9685198
R:Jones, S.M.; Palcic, M.M.; Scamam, C.H.; Smith, A.J.; Brown, D.E.; Dooley, D.M.;
Biochemistry 31, 12147-12154, 1992
A:Title: Identification of topaquinine and its consensus sequence in copper amine
A:Reference number: A44239; MID:93090746; PMID:11457410
A:Accession: C44239
A:Molecule type: Protein
A:Residues: 409-417, 'X', 419 <JAN>
A:Experimental source: seedling
C:Superfamily: amine oxidase (copper-containing)
C:Keywords: copper; glycoprotein; metalloprotein; oxidoreductase; quinoprotein; tol
F:1-25/Domain: signal sequence #status predicted <SIG>
F:126-674/Product: amine oxidase (copper-containing) #status predicted <MAT>
F:166-389,583/Binding site: carbonylurate (Asn) (covalent) #status predicted
F:382,467/Binding site: copper (His) #status predicted
F:412/Modified site: topaquinine (Tyr) #status experimental

Query Match 8.0%; Score 320; DB 2; Length 674;
Best Local Similarity 25.8%; Pred. No. 3.2e-17;
Matches 120; Conservative 66; Mismatches 189; Indels 90; Gaps 20;

Qy 282 SLKSPVPPGAPP--LQTFPGPRSVSGSRVASSLMTFSGLAGFSG-----PRIFPV- 333
Db 233 SKQSP-PGPKPOHSLTSHQPOGPGQINGHSVSMANMFHIGEDVRAIGVYSLASTYDLE 291
Qy 334 RFQGERLYEYSLOEALAYGNSPAAATTTRYVDG-GFGMKYTTPLTRGVDCPYLATYV 392
Db 292 KHKSRRLVLYKGISLFLPYDPTIEFFYKTFDSDGEGFGSLSTSLIPNRCCPPHAQFI 351
Qy 393 DMHFLLESQAPKTRIDAFVCEONOGPLRLRHSDLSHYFGGL-----AETLVVY 443
Db 352 DTIVHASANGTPIILKNALCVFEQ-----YGNIMMRHTENGIPNESIESRTEVNLI 403
Qy 444 RSMSTLLTYDVMVDFVHPSCAIEIRFYATGYISSAFLGAGK-----YGNQVSE 494
Db 404 RTIVTVGNDVYIDMEFASGSGIKPSIALSGILE---IKGTINKRDEIKEDLHGKLYSA 460
Qy 495 HTLGVTHHSNAHFVLDVAGLENNVMAEDNVFVMAVPWSPEHOLRLOVTR 547
Db 461 NSIGTYHDHFYIYTDLIDGTGTHNSFEKSLTIVRKDDSSKRSKRWTEGTQAK----- 515
Qy 548 KLEMEQAAFLVGSATPRYLILASNHSGNKGHPRGYRIOMLSFAGEPL-----PONS 601

```
Db 516 ----TESDAKITIGLAPALVYVNNIKTAVGNEVGYRL-IPAIPAHLITEDDYPO--- 567
QY 602 MARGESMERIOLAVTQKKEEPPSSSVFNQNDPMAP--TVDES---DFI-----NNETIA 651
Db 568 -IRG-AFTNYVWVY-----AYNRTEKNAGGLYVDHSRGDDTLAVWTKQNRREIV 614
QY 652 GKDLVAVWTAGFLHPHADIPTVTVGNGVGFILRPFYNEFDEDP 696
Db 615 NKDLYMHVHVGIIHVPAQDEPIMPPL--STSFELRPTNFEERNP 657
```

Search completed: May 20, 2003, 12:05:04
Job time : 23 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:41:09 ; Search time 17 Seconds
(without alignments)
1815.199 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763
Sequence: 1 LVCVLLVGRGSDGGEPSQLP.....QNAACAPDLPAFSGGFSHN 744

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4010	100.0	763	1 AOC3_HUMAN	016853 homo sapien
2	3404	84.9	765	1 AOC3_MOUSE	070423 mus musculu
3	3343	83.4	762	1 AOC3_BOVIN	046406 bos taurus
4	3331	83.1	762	1 AOC3_BOVIN	029437 bos taurus
5	2565	64.0	756	1 AOC2_HUMAN	075106 homo sapien
6	1398.5	34.9	746	1 AOB_RAT	P36633 rattus norv
7	1394	34.8	751	1 AOB_HUMAN	P19601 homo sapien
8	1354	33.8	751	1 AOC3_RAT	008590 rattus norv
9	413	10.3	755	1 AOC_KLEAE	P49250 klebsiella
10	397.5	9.9	757	1 AOC_HUMAN	P46883 escherichia
11	377	9.4	683	1 AOC_HUMAN	059118 arthrobacte
12	363	9.1	648	1 AOC2_ARTSI	007121 arthrobacte
13	363	9.1	648	1 AOC2_ARTSI	007123 arthrobacte
14	340	8.5	638	1 PAOX_ARTGO	P46881 arthrobacte
15	320	8.0	674	1 AOC_PEA	P43077 pisum sativ
16	311.5	7.8	666	1 AOC_LENCU	P49252 lens culina
17	310	7.7	671	1 AOC_LENCU	P49252 lens culina
18	295	7.4	692	1 AOC_PICAN	P12807 picella fas
19	111	2.8	1322	1 PUR4_XYLEA	090416 xylella fas
20	103	2.6	2224	1 COX2_HUMAN	P12859 homo sapien
21	102.5	2.6	591	1 COX2_HUMAN	P98000 bradyrhizob
22	101.5	2.5	736	1 DVL2_MOUSE	060838 mus musculu
23	101.5	2.5	1520	1 ACDP_ECOLI	046837 escherichia
24	101	2.5	435	1 KICH_MOUSE	054804 mus musculu
25	100.5	2.5	1237	1 KICH_MOUSE	P32004 homo sapien
26	100	2.5	736	1 DVL2_HUMAN	O14641 homo sapien
27	98.5	2.5	881	1 NIAL_PHAVU	P39665 phaseolus v
28	98.5	2.5	1217	1 S3B3_HUMAN	P39665 phaseolus v
29	98	2.4	2212	1 RPL1_EBOZM	005118 ebola virus
30	97	2.4	733	1 HEXA_GLADI	O17127 blaberus di
31	96.5	2.4	1137	1 RIR1_HSA11	P08543 herpes simp
32	96	2.4	421	1 SYS_THERH	P34945 thermus the
33	95.5	2.4	553	1 C862_ARATH	O23066 arabidopsis

34	95.5	2.4	985	1 AGU1_ASPNG	P56526 aspergillus
35	95.5	2.4	2175	1 POLG_BOVEV	P12315 bovine ente
36	95	2.4	2258	1 PA5_PIG	O991P1 sus scrofa
37	94.5	2.4	901	1 VEF_GYPU	P41723 pseudolatia
38	94.5	2.4	1522	1 PST1_SCHPO	O09750 schizosacch
39	94.5	2.4	2555	1 PPS3_BACSU	P39847 bacillus su
40	94	2.3	492	1 SYM_YEAST	P48527 saccharomyc
41	94	2.3	716	1 DVL3_HUMAN	O92897 homo sapien
42	94	2.3	716	1 DVL3_MOUSE	O61062 mus musculu
43	93.5	2.3	878	1 FIMD_ECOLI	P30130 escherichia
44	93	2.3	697	1 YN26_MYCTU	P71886 mycobacteri
45	93	2.3	958	1 MML1_MYCTU	P95211 mycobacteri

ALIGNMENTS

```

RESULT 1
AOC3_HUMAN          STANDARD:      PRT:      763 AA.
AC  016853;
DT  01-NOV-1997 (rel. 35, Created)
DT  01-NOV-1997 (rel. 35, Last sequence update)
DT  15-JUN-2002 (rel. 41, Last annotation update)
DE  Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-
DE  1) (VAP-1) (HPAO).
GN  AOC3 OR VAP1.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Placenta;
RX  MEDLINE=97128319; PubMed=8972912;
RA  Zhang X., McIntire W.S.;
RT  "Cloning and sequencing of a copper-containing, topaquinine-
RT  containing monamine oxidase from human placenta.";
RL  Gene 179:279-286(1996).
RN  [2]
RP  SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC  TISSUE=Lung;
RX  MEDLINE=98317014; PubMed=9653080;
RA  Smith D.J., Salini M., Bono P., Hellman J., Leu T., Jalkanen S.;
RT  "Cloning of vascular adhesion protein 1 reveals a novel
RT  multifunctional adhesion molecule.";
RL  J. Exp. Med. 188:17-27(1998).
CC  - FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE
CC  RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO
CC  PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
CC  INDEPENDENT FASHION. HAS A MONAMINE OXIDASE ACTIVITY.
CC  - CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC  H(2)O(2).
CC  - COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC  - SUBUNIT: HOMODIMER.
CC  - SUBCELLULAR LOCATION: Type II membrane protein.
CC  - TISSUE SPECIFICITY: MOST STRONGLY EXPRESSED ON THE HIGH
CC  ENDOTHELIAL VENULES OF PERIPHERAL LYMPH NODES AND ON HEPATIC
CC  ENDOTHELIAL.
CC  - PTM: Topaquinine (TPQ) is generated by copper-dependent
CC  autooxidation of a specific tyrosyl residue (By similarity).
CC  - PTH: N- AND O-GLYCOSYLATED.
CC  - SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC  - SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
DR  EMBL: U39447; AAC50919.1; -

```

```

Dh 560 QRLVTRLELMEEQOAAFLVGSAPRYLYLASNSNKKGHPRGRIOMLSPFAGEPLDPS 619
Qy 601 SMAGFSMERQQLAVTORKEEPPSSSVENONNDMAPIVDSPTINNTIGKDLVAVT 660
Db 620 SMAGFSMERQQLAVTORKEEPPSSSVENONNDMAPIVDSPTINNTIGKDLVAVT 679
Qy 661 AGFLHPHAEIDIPMTVTVGNGVGFFLRPNFDEDEPSFYSDSYFRGDODAGACEVNP 720
Db 680 AGFLHPHAEIDIPMTVTVGNGVGFFLRPNFDEDEPSFYSDSYFRGDODAGACEVNP 739
Qy 721 ACPDQAAACAPDLPAFSHGGSFN 744
Db 740 ACPDQAAACAPDLPAFSHGGSFN 763

RESULT 2
AOC3.MOUSE
ID AOC3.MOUSE STANDARD: PRT: 765 AA.
AC 070423;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1).
EN AOC3 OR VAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NX NCBI_Taxid=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/c, and 129/SvJ.
RX MEDLINE=98414290; PubMed=9743358;
RA Bono P., Salm M., Smith D.J., Leppanen I., Horelli-Kuitunen N., Palotie A., Jalkanen S.;
RT "Isolation, structural adhesion characterization, and chromosomal mapping of the mouse vascular adhesion protein-1 gene and promoter.";
RL J. Immunol. 161:2953-2960(1998).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-INDEPENDENT FASHION. HAS A MONOMINE OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) + H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent autooxidation of a specific tyrosyl residue (By similarity).
CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on ways to use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@isb-sdb.ch).
CC
DR EMBL, AF054831; AAC23747.1; -
DR EMBL, AF078705; AAC35839.1; -
DR GDB: MGI:1306797; AOC3.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid. 1.
DR Pfam: PF02127; Cu_amine_oxidn2. 1.
DR Pfam: PF02728; Cu_amine_oxidn3. 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER AMINE_OXID.1; 1.
DR PROSITE: PS01165; COPPER AMINE_OXID.2; 1.
DR Oxidoreductase; Copper; TPO; Glycoprotein; Transmembrane;
KW Signal anchor; Cell adhesion; Metal-binding
FT DOMAIN 1 6
CTOPLASMIC (POTENTIAL).

```

```

FT TRANSMEM 7 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 28 765 (POTENTIAL).
FT MOD_RES 471 471 EXTRACELLULAR (POTENTIAL).
FT METAL 520 520 TOPOQUINONE (BY SIMILARITY).
FT METAL 522 522 COPPER (POTENTIAL).
FT METAL 522 522 COPPER (POTENTIAL).
FT METAL 684 684 COPPER (POTENTIAL).
FT BINDING 673 673 AMILORIDE (BY SIMILARITY).
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 659 659 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 765 AA; 84533 MW; 7489DE673DBB44D CRC64;

Query Match 84.9%; Score 3404; DB 1; Length 765;
Best Local Similarity 83.2%; Pred. No. 1.3e-259;
Matches 618; Conservative 61; Mismatches 64; Indels 0; Gaps 0;

QY 1 LVCVLLVGRGDSGEPSQLEPHCSVSPSAQPMTHPGQSOLFADLSREELTAVMKFLTQRL 60
DB 20 LVCVLLVGRGDSGEPSQLEPHCSVSPSAQPMTHPGQSOLFADLSREELTAVMKFLTQRL 79
QY 61 GPGVLAQAARPSDNCVFSEYLOLPRKAAALHLDRGSPPRAREALATVFFGRPOPQNV 120
DB 80 GPGVLAQAARPSDNCVFSEYLOLPRKAAALHLDRGSPPRAREALATVFFGRPOPQNV 139
QY 121 ELVVGPLPHSPYMRDVTVERHGRLPRHRRPVLFQYELDIIDOMIFNELFOASGLHHC 180
DB 140 ELVVGPLPHSPYMRDVTVERHGRLPRHRRPVLFQYELDIIDOMIFNELFOASGLHHC 199
QY 181 FYKHGRNLTMTTAPRGLOSGDRATWFGDYLYNISGAFGLHVGELLYNHNKALDPRM 240
DB 200 FYKHGRNLTMTTAPRGLOSGDRATWFGDYLYNISGAFGLHVGELLYNHNKALDPRM 259
QY 241 TIKVYVQGRYVDSLAQLAQFENAGLVNVLIDNKGSGWSLSPRPPAPRPLQYPO 300
DB 260 TIKVYVQGRYVDSLAQLAQFENAGLVNVLIDNKGSGWSLSPRPPAPRPLQYPO 319
QY 301 GPRFVSQGSVASSLMTFSGLGAFSGPRIFDVRFQGERLYVEYSIQLEALATYGNPAA 360
DB 320 GPRFVSQGSVASSLMTFSGLGAFSGPRIFDVRFQGERLYVEYSIQLEALATYGNPAA 379
QY 361 MTRRYVDGFGKMKYTPPLTRGYDCRYLATYVDMHFLLESQAPKTRDAFCVEQNGLP 420
DB 380 MTRRYVDGFGKMKYTPPLTRGYDCRYLATYVDMHFLLESQAPKTRDAFCVEQNGLP 439
QY 421 LRHHNSDLYHYFGGLAETLVYRSMSTLLNYDMYVMDTVPHPSGATIRRYATCYISSAF 480
DB 440 LRHHNSDLYHYFGGLAETLVYRSMSTLLNYDMYVMDTVPHPSGATIRRYATCYISSAF 499
QY 481 LFGATKYGNQVSEHTLGTVHTSHAFKVDLDVAGLENNMAEDMVEVPMAVSPSHOL 540
DB 500 LFGATKYGNQVSEHTLGTVHTSHAFKVDLDVAGLENNMAEDMVEVPMAVSPSHOL 559
QY 541 ORLQVTRKLLLEBQAAFLVGSATPRYLALASNNKMGHPRGRYRIOMLSFAGEPLDONS 600
DB 560 ORLQVTRKLLLEBQAAFLVGSATPRYLALASNNKMGHPRGRYRIOMLSFAGEPLDONS 619
QY 601 SMARGFSEMERVYOLAVTORKEEPSSSVFNONDPMAATVDFSDPINNETAGKULVAVMT 660
DB 620 SMARGFSEMERVYOLAVTORKEEPSSSVFNONDPMAATVDFSDPINNETAGKULVAVMT 679
QY 661 AGELTHPHADIDPNTVYVNGVGFELRYNFEFDEDPSTFYADSIFYEGDDAGACEVNP 720
DB 680 AGELTHPHADIDPNTVYVNGVGFELRYNFEFDEDPSTFYADSIFYEGDDAGACEVNP 739
QY 721 ACLEPAAACAPDLPAFSGHGF 743
DB 740 ACLEPAAACAPDLPAFSGHGF 762

```

```

RESULT 3
ACQY_BOVIN
ID ACQY_BOVIN STANDARD; PRT; 762 AA.
AC 046406;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase, lung isozyme precursor (EC 1.4.3.6) (Amine
DE oxidase [copper-containing]) (BOLA0).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bos.
OX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=98151264; PubMed=9492300;
RA Hoggadell E.V.S., Houen G., Botre W., Bundgaard J.R., Larsson L.-I.,
RA Vuust J.;
RT "Structure and tissue-specific expression of genes encoding bovine
RT copper amine oxidases."
RL Eur. J. Biochem. 251:320-328(1998).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit (By
CC similarity).
CC -1- ENZYME REGULATION: INHIBITED BY AMILORIDE IN A COMPETITIVE MANNER
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Extracellular (By similarity).
CC -1- TISSUE SPECIFICITY: LUNG, SPLEEN, HEART, KIDNEY.
CC -1- PTM: Topaquinone (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (By similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPOQUINONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC
DR EMBL: Y15774; CAA75776.1;
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF011179; Cu_amine_oxid. 1.
DR Pfam: PF02727; Cu_amine_oxid2. 1.
DR Pfam: PF02728; Cu_amine_oxid3. 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID. 1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID. 2; 1.
DR Oxidoreductase; Copper; TPQ; Glycoprotein; Signal.
KW CHAIN 1 762
FT SIGNAL 17 762
FT MOD_RES 470 470 TOPOQUINONE (BY SIMILARITY).
FT METAL 519 519 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT METAL 683 683 COPPER (POTENTIAL).
FT BINDING 672 672 AMILORIDE (BY SIMILARITY).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 762 AA; 84883 MW; B43D04776744A2 CRC64;

Query Match 83.4%; Score 3343; DB 1; Length 762;
Best Local Similarity 82.8%; Pred. No. 8.1e-255;
Matches 619; Conservative 44; Mismatches 75; Indels 10; Gaps 2;

QY 4 VLVGRL--GGDGE-----PSQLPHCSVSPSAQPMTHPGQSOLFADLSREELTAVM 53
DB 12 VLVGREGGVGSEGVGKCHPSLPDPCPSRSPSDQPMTHPGQSOLFADLSREELTAVM 71

```

```

QY 54 RFLTORLPGVLVDAQAQAPSPDNCVSVLEQLPPKAAALAHLDKSPPPARALAIYFEGR 113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72 SFLUQKGRPLVDAQAQAPSPDNCVSVLEQLPPKAAALAHLDKSPPPARALAIYFEGG 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 QOPPNVSELVGLPHPSYMRDVTVERHGGRLPYHRRPVLFQEYLDIDQMFNRELPOAS 173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 132 QOPPNVSELVGLPHPSYMRDVTVERHGGRLPYHRRPVLFQEYLDIDQMFNRELPOAA 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 GLHHCCEYKRRGNLYMTTAPRGLOSGDRATWFGLYYNSGAGFLHNGELLYVNHK 233
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 192 GVLHHCCEYKRRGNLYMTTAPRGLOSGDRATWFGLYYNSGAGFLHNGELLYVNHK 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 234 ALDPARTIOKVFQYGRYYSLAOLEAOFEGALVNVLLIPDNGGSGWSLSPPPGPAP 293
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 252 ALDPADMTIOKVFQYGRYYSLEVLQLEOFEGAGRVNVVYIPDNGGSGWSLSPPPGPPTP 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 294 PLOFYPOGPRFSYVQSGSVASSLWTFSGLAFSGPRIFDVRFOGERLYEISLOEALAIY 353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 312 PLOFYPOGPRFSYVQSGSVASSLWTFSGLAFSGPRIFDVRFOGERLYEISLOEALAIY 371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 354 GGNPRAMTRRYVDGGRGMRKYTTPLTRGYDCPYLATYVDMHFLLESQAPRTLDARCVF 413
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 372 GGNPRAMTRRYVDGGRGMRKYTTPLTRGYDCPYLATYVDMHFLLESQAPRTLDARCVF 431
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 414 EONQGLPLRRHSDLYSHYFGLAETVLYVRSMTLLNDYDVTVEHPSGAIEIRFYAT 473
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 432 EONQGLPLRRHSDLYSHYFGLAETVLYVRSMTLLNDYDVTVEHPSGAIEIRFYAT 491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 474 GYSSAFLFGATGKYGNVSEHLYGVYHSHAFKVDLDVAGLENNWVAEDMVEFPAAPV 533
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 492 GYSSAFLFGATGKYGNVSEHLYGVYHSHAFKVDLDVAGLENNWVAEDMVEFPAAPV 551
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 534 WSPHOLORLOVNFRLKLEMEBOAFLVGSAPRYTLASNSNGHGRGRIGIOMLSRAG 593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 552 WSPHOLORLOVNFRLKLEMEBOAFLVGSAPRYTLASNSNGHGRGRIGIOMLSRAG 611
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 594 EPLPONSMAKGFSEWERYQOLAVTORKEEPPSSSVFNQNDPMAVTFVSDFINNETIAGK 653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 612 RPLPONSSTERALISWGRYQOLAVTORKEEPPSSSVFNQNDPMAVTFVSDFINNETIAGK 671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 654 DLVAVNTAGLHPIHADIPRTVTYVNGVGFLLPPIYFEDDPSFYADSIYFRGDDAG 713
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 672 DLVAVNTAGLHPIHADIPRTVTYVNGVGFLLPPIYFEDDPSFYADSIYFRGDDAG 731
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 714 ACEVNPPLACLPQAAACAPDLPFASHGCF 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 732 SCEINPLACLPQAAACAPDLPFASHGCF 759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 4 AOCX_BOVIN STANDARD; PRT; 762 AA.

```

AC 029437;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase, liver isozyme precursor (EC 1.4.3.6) (Amine
OS oxidase [copper-containing]) (Serum amine oxidase) (SAO).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RA MEDLINE=9419366; PubMed=8144587;
RA Burillogame A.L., Smith A.J., Cai D., Kilman J.P.;
RT "Primary structures for a mammalian cellular and serum copper amine
RT oxidase."
RL J. Biol. Chem. 269:9926-9932(1994).

```

```

RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=9309748; PubMed=1457410;
RA Jones S.M., Palcic M.M., Scaman C.H., Smith A.J., Brown D.E.,
RA Dooley D.M., Mure M., Kilman J.P.;
RT "Identification of topaquinine and its consensus sequence in copper
RT amine oxidases."
RL Biochemistry 31:12147-12154(1992).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- ENZYME REGULATION: INHIBITED BY AMILORIDE IN A COMPETITIVE MANNER.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Liver.
CC -1- PM: Topaquinine (TPO) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (by similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S69583; AAB30397.1; -.
DR EMBL: L27218; AAB30525.1; -.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUNDOXIDSE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
KW Oxidoreductase; Copper; TPO; Glycoprotein; Signal; Metal-binding.
FT SIGNAL 1
FT CHAIN 17 762 COPPER AMINE OXIDASE, LIVER ISOZYME.
FT MOD_RES 470 470 TOPAQUINONE (BY SIMILARITY).
FT METAL 519 519 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT METAL 683 683 COPPER (POTENTIAL).
FT BINDING 672 672 AMILORIDE (BY SIMILARITY).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 762 AA; 84756 MW; AA959771360295FE CRC64;

```

Query Match 83.1%; Score 3331; DB 1; Length 762;
 Best Local Similarity 81.4%; Pred. No. 7.1e-254;
 Matches 609; Conservative 59; Mismatches 70; Indels 10; Gaps 2;

```

QY 4 VLVYGR--GGDGGF-----PSQLPHCPSPVSSAQPWTHPGQSOLFADLSREELTAVM 53
    ||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12 LLVNKGREGGVSGSEGVGKQCHPSLPKPCPSRSPSDPMTHPDQSLFADLSREELTAVM 71
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54 RFLTORLPGVLVDAQAQAPSPDNCVSVLEQLPPKAAALAHLDKSPPPARALAIYFEGR 113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72 SFLUQKGRPLVDAQAQAPSPDNCVSVLEQLPPKAAALAHLDKSPPPARALAIYFEGG 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 QOPPNVSELVGLPHPSYMRDVTVERHGGRLPYHRRPVLFQEYLDIDQMFNRELPOAS 173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 132 QOPPNVSELVGLPHPSYMRDVTVERHGGRLPYHRRPVLFQEYLDIDQMFNRELPOAA 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 GLHHCCEYKRRGNLYMTTAPRGLOSGDRATWFGLYYNSGAGFLHNGELLYVNHK 233
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 192 GVLHHCCEYKRRGNLYMTTAPRGLOSGDRATWFGLYYNSGAGFLHNGELLYVNHK 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 234 ALDPARTIOKVFQYGRYYSLAOLEAOFEGALVNVLLIPDNGGSGWSLSPPPGPAP 293
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 252 ALDPADMTIOKVFQYGRYYSLEVLQLEOFEGAGRVNVVYIPDNGGSGWSLSPPPGPPTP 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 294 PLOFYPOGPRFSYVQSGSVASSLWTFSGLAFSGPRIFDVRFOGERLYEISLOEALAIY 353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 312 PLOFHPGGRFVSQGNRASSLMTFSFGAGFSGPRVDFVROGGERFLAYEISLOEGAY 371
QY 354 GGNPAAMTTRRYVGGFGMGKYYTTPTRGCDPYLATYVWHFLESQAPKTRDAFCVF 413
DB 372 GGNTPAAMLTRYMDSGFGMGYFATPLIRGDCPYLATYVDMHFVYESPKTLHDAFCVF 431
QY 414 EONOGILPLRRHSDLSHFHFGGLAEVLYVRSKSTILANDYMDVYFHSGAIEIFAT 473
DB 432 EONKGLPLRRHSDLSHFHFGVGAQVTLFRSVSTMLANDYMDVYFNGAIEVYLAAT 491
QY 474 GTISSAFELGATGKGNVSEHTLGTVTHSAHFKVDLVAGENYMAEDMAYFVMAVP 533
DB 492 GTISSAFELGAAARRIGNVGEHTLGTVTHSAHYKTDVGLGLENYMAEDMAYFVTAIP 551
QY 534 WSPHQLOQLQVTRKLLMEQAAFLVGSATPRYLYLASNHSKMGHPRGYRIQMLSPAG 593
DB 552 WSPHQIQLQVTRKQLETEQAAFLVGSATPRYLYLASKQSKMGHPRGYRIQVTSFAG 611
QY 594 EPLPQSSMARGFSMRYGLATYQKKEEPPSSSVNODPAPTYVDFDFTINNETIACK 653
DB 612 GMPQPSMERAFSMGRYGLATYQKKEEPPSSSVNODPAPTYVDFDFTINNETIACK 671
QY 654 DLVAVWTAGFLHPRHEDIPNTVGVNGVGFELRPYFDEDEPSFYSADSIYRGQDAG 713
DB 672 DLVAVWTAGFLHPRHEDIPNTVGVNGVGFELRPYFDEDEPSFYSADSIYRGQDAG 731
QY 714 ACEVNPACLPQAAACAPDLPAFSGGF 741
DB 732 SCEINPLACLPAATCAPLPVYSHGX 759

RESULT 5
AOC2_HUMAN
ID AOC2_HUMAN STANDARD; PRT; 756 AA.
AC 075106; 075105; 000120; 09UNY0;
AD 15-JUL-1999 (Rel. 38, Created)
AD 15-JUL-1999 (Rel. 38, Last sequence update)
AD 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retina-specific copper amine oxidase precursor (EC 1.4.3.6) (RAO)
DE (Amine oxidase [copper-containing]).
GN AOC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=97237047; PubMed=9119395;
RA Imanura Y., Kubota R., Wang Y., Asakawa S., Kudoh J., Mashima Y.,
RA Oguchi Y., Shimizu N.;
RA "Human retina-specific amine oxidase (RAO): cDNA cloning, tissue
RT expression, and chromosomal mapping.";
RL Genomics 40:277-283(1997).
RN 121
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=98390194; PubMed=9722954;
RA Imanura Y., Noda S., Mashima Y., Kudoh J., Oguchi Y., Shimizu N.;
RA "Human retina-specific amine oxidase: genomic structure of the gene
RT retinal", alternatively spliced variant, and mRNA expression in
RL Genomics 51:293-298(1998).
RN 131
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Zhang X., McIntire W.S.;
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE A CRITICAL MODULATOR OF SIGNAL TRANSMISSION IN
CC RETINA, POSSIBLY BY DEGRADING THE BIOGENIC AMINES DOPAMINE,
CC HISTAMINE, AND PUTRESCINE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

```

```

CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaguinone per subunit.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: RETINAL-SPECIFIC.
CC -1- PTM: Topaguinone (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: AB012943; BAA32590.1; -.
DR EMBL: AB012942; BAA32590.1; JOINED.
DR EMBL: AB012943; BAA32589.1; -.
DR EMBL: AB012942; BAA32589.1; JOINED.
DR EMBL: D88213; BAA19001.1; -.
DR EMBL: AF081363; AAD39345.1; -.
DR Genew: HGNC:549; AOC2.
DR MIM: 602268; -.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF02179; Cu_amine_oxid. 1.
DR Pfam: PF02177; Cu_amine_oxid2; 1.
DR Pfam: PF02178; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER AMINE OXID-1; 1.
DR PROSITE: PS01165; COPPER AMINE OXID-2; 1.
KW Oxioreductase; Copper; TPQ; Alternative splicing; Glycoprotein;
KW Signal; Metal-binding.
FT SIGNAL 1 32
FT CHAIN 33 756
FT MOD_RES 465 465
FT METAL 516 516
FT METAL 518 518
FT METAL 680 680
FT BINDING 669 669
FT CARBOHYD 133 133
FT CARBOHYD 198 198
FT CARBOHYD 226 226
FT CARBOHYD 588 588
FT CARBOHYD 662 662
FT VARSPLIC 599 625
FT CONFLICT 181 181
FT CONFLICT 215 218
FT CONFLICT 221 222
FT CONFLICT 610 610
SQ SEQUENCE 756 AA; 83793 MW; 4E3B7317E6DAC66 CRC64;

Query Match 64.0%; Score 2565; DB 1; Length 756;
Best local Similarity 65.3%; Pred. No. 1,le-193;
Matches 485; Conservative 88; Mismatches 162; Indels 8; Gaps 3;

QY 1 LVCVLLVGRGCGDEPSQLPHCPVSFSAQPMWTHPGOSQLFADLSREELTAVNRFLTQRL 60
DB 20 LAVVLLTSPGSGS----SQPHCPVSFSAQPMWTHPGOSQLFADLSREELTAVNRFLTQRL 75
QY 61 GPGVLDAQAQRPSDNCVFSVELDLPKAAALAHLDKRSPPARBAALAIYFSGQPOPNVS 120
DB 76 GPGVLDAQAQAQPSDNCIFSVYELDLPKAAALAHLDKRSPPARBAALAIYFSGQPOPNVS 135
QY 121 ELVVGPLPHPSYKRDVTVRHGGPLPYHRRPVLEQYELDIDOMIFNELPOASGLHHC 180
DB 136 ELVVGPLPHPSYKRDVTVRHGGPLPYHRRPVLEQYELDIDOMIFNELPOASGLHHC 194
QY 181 FYKRRGNLTWTAPRGLOSGRATWFGLYYNSAGFELHVGLELLVNHKRALDPARW 240
DB 195 -FMYNGSTLAAVHATPRGLNSRETTMMALYHNISGVGLFHLHPVGLLELLDHALDPAHW 253

```



```

DB 496 LQTHLLGNHTLVHVRVDMDVAGTKNSFQTLTKMLKLENTLNPWSPSHSLVQPTLEOTQYS 555
OY 552 MEEQAAFLVGSATPRYLTLASNSHNGHGRGYRIOMLSPAGEPLPONSSMAROFSEWERY 611
DB 556 QEQHQAAPFRGQGLPKYLLPSSPOKNCMGHRSTRLOIHSNAEYVLPPGQNEERAVTYARY 615
OY 612 QLAVTORKEEPESSSVFVNDPMAPTVDFSDFI-NNETIAGKDLVAVMTAGFLHPHAE 670
DB 616 PLAVTKYRESERSSSLYNQNDPMDPVVEEFLRNENIETEDLVAVMTVGLHPHSE 675
OY 671 DIANTVTGNGVGFLLRPVYFEDDEPSFYGSADSIYFGDDAGACEVNPPLACLPQAACA 730
DB 676 DVNPTATPGNSVGFLLRPVYFEDDEPSLASRDVTVV-WPDCKGLNRVO--RWIPEDRRCL 732
OY 731 PDLPAFSGHG 740
DB 733 VS-PFSSYNG 741

RESULT 7
ABP_HUMAN
ID ABP_HUMAN STANDARD: PRT: 751 AA.
AC P19801: Q16683; Q16684;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Amloride-sensitive amine oxidase [copper-containing] precursor
DE (EC 1.4.3.6) (Diamine oxidase) (DAO) (Amloride-binding protein)
DE (ABP) (Histaminase).
GN ABP1 OR DAO1 OR AOC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-94237856; Pubmed-8182053;
RX Chassande O., Renard S., Barbry P., Lazdunski M.;
RT "The human gene for diamine oxidase, an amloride binding protein.
RT Molecular cloning, sequencing, and characterization of the promoter.";
RL J. Biol. Chem. 269:14484-14489(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE-Placenta; Pubmed-8595053;
RA Zhang X., Kim J., McIntire W.S.;
RT "cDNA sequences of variant forms of human placenta diamine oxidase.";
RL Biochem. Genet. 33:261-268(1995).
RN [3]
RP PRELIMINARY SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE-Kidney;
RA MEDLINE-91017502; Pubmed-2217167;
RX Barbry P., Champe M., Chassande O., Munemitsu S., Champigny G.,
RA Lingueglia E., Maes P., Frelin C., Tatar A., Ullrich A.,
RA Lazdunski M.;
RT "Human kidney amloride-binding protein: cDNA structure and functional
RT expression.";
RN Proc. Natl. Acad. Sci. U.S.A. 87:7347-7351(1990).
RN [4]
RP SEQUENCE OF 20-39, AND CHARACTERIZATION.
RX TISSUE-Placenta;
RX MEDLINE-94193685; Pubmed-8144586;
RA Novotny W.F., Chassande O., Baker M., Lazdunski M., Barbry P.;
RT "Diamine oxidase is the amloride-binding protein and is inhibited by
RT amloride analogues.";
RL J. Biol. Chem. 269:9921-9925(1994).
CC -1- FUNCTION: CATALYZES THE DEGRADATION OF COMPOUNDS SUCH AS
CC PUTRESCINE, HISTAMINE, SPERMINE, AND SPERMIDINE, SUBSTANCES
CC INVOLVED IN ALLERGIC AND IMMUNE RESPONSES, CELL PROLIFERATION,
CC TISSUE DIFFERENTIATION, TUMOR FORMATION, AND POSSIBLY APOPTOSIS.
CC PLACENTAL DAO IS THOUGHT TO PLAY A ROLE IN THE REGULATION OF THE
CC FEMALE REPRODUCTIVE FUNCTION.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

```

```
DB 197 RRSMLIIQRYVE--GYFLHPTGLLELVHDSSTDAHGHWAVEQWYNGKFGSPDEELARKYA 254
QY 264 AGLVNVLIPNGTGGMSLSKSPVR-----GAP-----PLOGYPPGPPSPVQ 308
DB 255 DGEVDVNVLEPPLPGGKHDSSTEEPLSSSHKPRGDFSPPIHVSQPRLVQPPFRFRLEG 314
QY 309 SRVASSLTFFSGGAFSGPRIFDVRFOGGERLYVEISLOELALYIGNSPAMATTRYVDG 368
DB 315 NAVLYGSGSFARLSSSGLOLVANHFEGGERIAYEVSOEVALYGGHFRAGOMOKRYLDV 374
QY 369 GFGMKRYTPTLRGVDCPYLATYVDMHFLLESQAPKTRDAFCVEQONQGLPLRRHSD 428
DB 375 GNGLSQVYHEALPAGIDCPETATFLDTFHYDADDVNHYPALCLPEMPFGVPLRRHNSN 434
QY 429 YS---HYEGGLAEPLVYRSMSTLLNLYVMDTVPHPSGALIEIRYANGYISALFGAT 485
DB 435 FKGGFNFAAGLKGOLVLTSTYNTYIDFETYPNGVAKAKHAGYVATYTPREG 494
QY 486 GKYNQVSEHTLGIVTHSAHFYDLADYAGLENNYMAEDMVFVPAVWSPSEHQLQRY 545
DB 495 LRHGRLHTHLIGNHTLVHYRVDLDVAGTKNSFQTLQMKENTNTNPMSPRRVVOPTL 554
QY 546 TRKLEMEQAAFLVGSATPRILYLAHNSKMGHPRGYRIMSLFAGEPLPONSMAWG 605
DB 555 EOTOYSWRQAAFRFKRLPYLLFTSPQENPMGHRKSRRLQIHSMAQOVLPFGMOEQA 614
QY 606 FSWERYOLAVTORKEEPPSSSVFNQNDPMAPTVDSDFI--NNETIAGDIAVMTAGTL 664
DB 615 ITMAYPLAVTKYRSELCSSITIHQNDPMHPVVEGFLHNENIENDLVAWTVGTL 674
QY 665 HIPHAEIPNTVYNGVGFELRPYFDEDEPSYSADSIYFRGDQDAGACEVNPACLP 724
DB 675 HIPHSEDIPNATPGNSVGLFRPFPEDPSLASRDTVLY-WRDNGPNVYQ--RWLP 731
QY 725 QAACAPDLPAFSHG 740
DB 732 EDRDCSMP-PPFSYNG 746

RESULT 8
AOC3_RAT STANDARD: PRT: 321 AA.
AC 008590:
AT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1) (Vp97) (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-Sprague-Dawley; TISSUE-Adipocyte;
RX MEDLINE=97238878; PubMed=9083076;
RA Morris N.J., Ducret A., Aebersold R., Ross S.A., Keller S.R., Lienhard G.E.;
RT "Membrane amine oxidase cloning and identification as a major protein in the adipocyte plasma membrane."
RL J. Biol. Chem. 272:9388-9392(1997).
RN (2)
RP SEQUENCE OF 1-19.
RC TISSUE-Adipocyte;
RX MEDLINE=96010417; PubMed=8520629;
RA Jochen A., Given S., Hays J.;
RT "The major integral membrane glycoprotein in adipocytes is a novel 200-kDa heterodimer."
RL Mol. Membr. Biol. 12:277-281(1995).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
```

```
CC INDEPENDENT FASHION. HAS A MONAMINE OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) + H(2)O(2).
CC
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquone per subunit.
CC
CC -1- SUBUNIT: HOMODIMER.
CC
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC
CC -1- PM: Topaquone (TPQ) is generated by copper-dependent autooxidation of a specific tyrosyl residue (by similarity).
CC
CC -1- PM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or\_send\_an\_email\_to\_license@sib-sib.ch).
CC
CC EMBL: U72632; AAC53189.1; -.
CC InterPro: IPR000269; CuNH_oxidase.
CC Pfam: PF02727; Cu_amine_oxidn2; 1.
CC Pfam: PF02728; Cu_amine_oxidn3; 1.
CC PRINTS: PRO0766; CUDAOXIDASE.
CC PROSITE: PS01164; COPPER_AMINE_OXID_1; PARTIAL.
CC PROSITE: PS01165; COPPER_AMINE_OXID_2; PARTIAL.
CC Oxidoreductase; Copper; TPQ; Glycoprotein; Transmembrane;
CC Signal-anchor; Cell adhesion; Metal-binding.
CC INIT_MBT 0
CC DOMAIN 1 5
CC TRANSMEM 6 26
CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 231 231
CC CARBOHYD 293 293
CC NON_TER 321 321
CC SEQUENCE 321 AA; 35116 MM; 6B1F294253A1DE6F CRC64;

Query Match 33.8%; Score 1354; DB 1; Length 321;
Best Local Similarity 83.2%; Pred. NO. 5e-99;
Matches 252; Conservative 18; Mismatches 33; Indels 0; Gaps 0;

QY 1 LVCVLVGRGSGDGEPSOLPHCPSPVPSAQPWTHGSQLFADLSREELTAVMRELTQRL 60
DB 19 LVCVLVLAGSGGGSLQSPKHPSPVLPQVQPTHSGGQGFADLSPEELTAVMSLIHL 78
QY 61 GPGVLDAAQARPSDNCVSVELQLPKAAALAHLDKSGPPAREKALAVFFGRQPPVNS 120
DB 79 GGLVDAQAQARPSDNCVSVELQLPKAAALAHLDKSGPPAREKALAVFFGRQPPVNS 138
QY 121 ELVGPPLPSPSMRYVTERHRCGLPYRRRPVLPQETVLDIDMIFNRELPOASGLLHHC 180
DB 139 ELVGPPLPSPSMRYVTERHRCGLPYRRRPVLPQETVLDIDMIFNRELPOASGLLHHC 198
QY 181 FYKHGRNLYVTTPARGLQSGDRATWGLYYNISGAFELHVGLELVNHNKALDPAW 240
DB 199 FYKHGRNLYVTTPARGLQSGDRATWGLYYNISGAFELHVGLELVNHNKALDPAW 258
QY 241 TIQKVFYGGRIYDLSAQLEAOFEAGLVNVLIPDNGTGGMSLSKSPVPPGAPPLQTYPQ 300
DB 259 TIQKVFYGGRIYDLSAQLEAOFEAGLVNVLIPDNGTGGMSLSKSPVPPGAPPLQTYPQ 318
QY 301 GPR 303
DB 319 GPR 321

RESULT 9
AMO_KLEAF STANDARD: PRT: 755 AA.
ID AMO_KLEAF
AC P49250:
DT 01-FEB-1996 (Rel. 33, Created)
```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Copper amine oxidase precursor (EC 1.4.3.6) (Monamine oxidase)
 DE (Tyramine oxidase).
 OS MAOA OR TYNA.
 OS Klebsiella aerogenes.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.
 NCBI_TaxID=28451;
 RN [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
 RC STRAIN=M70;
 RX MEDLINE=92210491; PubMed=1556068;
 RA Sugino H., Sasaki M., Azakami H., Yamashita M., Murooka Y.;
 RT "A monamine-regulated Klebsiella aerogenes operon containing the
 RT monamine oxidase structural gene (maoA) and the maoc gene.";
 RL J. Bacteriol. 174:2485-2492(1992).
 CC -1- FUNCTION: ACTIVE ON TYRAMINE, TRYPTAMINE, BETA-PHENETHYLAMINE AND
 CC DOPAMINE.
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaguinone per subunit.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY TYRAMINE AND CATECHOLAMINES.
 CC -1- PTM: Topaguinone (TPO) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D10208; BAA01060.1; .
 DR HSSP: P46883; 10AC.
 DR InterPro: IPR000269; CUNH_Oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxidn3_1.
 DR Pfam: PF02728; Cu_amine_oxidn3_1.
 DR PROSITE: PS01164; COPPER AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER AMINE_OXID_2; 1.
 KM Oxidoreductase; Copper; TPO; Periplasmic; Signal; Metal-binding.
 FT SIGNAL 1 30
 FT CHAIN 31 755 COPPER AMINE OXIDASE.
 FT ACT_SITE 413 413 CATALYTIC BASE (BY SIMILARITY).
 FT MOD_RES 496 496 TOPAGUINONE (BY SIMILARITY).
 FT METAL 554 554 COPPER (POTENTIAL).
 FT METAL 556 556 COPPER (POTENTIAL).
 FT METAL 719 719 COPPER (POTENTIAL).
 SEQUENCE 755 AA: 83576 MW: 785552283CD93EFF CRC64;
 Query Match 10.3%; Score 413; DB 1; Length 755;
 Best Local Similarity 24.1%; Pred. No. 1.5e-24;
 Matches 188; Conservative 101; Mismatches 284; Indels 206; Gaps 32;
 QY 14 GEPGSLPHCPSPVSPSAQPT-----HPGQSOLF-----DLSREELTAVMRF 55
 DB 82 GSKSLP-VPVYMKKGAAYSDTFINDVFGSGDFTGYEKRPHPLNSLSAIEISKAVT- 139
 QY 56 LTORLGPGLYDAQAARPSNCFVSVELQLPKRAALAHNDROGSRPPA-REALAIYVFGQ 114
 DB 140 -----IVKAPFQPTPTRETEISLHPRKAAVAFAALQGPVADPRTADVMDGK- 190
 QY 115 PPGNSLVYGLPHPSVYRDVTE--RHGGPLPYHRPFLFOEYLDIDIMFRELPAQS 173
 DB 191 ---HYEAVV-DLQNKKILSWPIKGAH-----MVLDDFVSQNTI----- 229
 QY 174 GLHHCFFYKHGRNLVTMTAPRGLOSGDRATWFLGYLNYISGAGFFLHVGLLELVNKH 233
 DB 230 -----NTSEFE-----EVLKKHG 243

QY 234 ALDPARWTQKV---FYGRYDLSLAQLEAGPEAGLVNVLIPDNKGGSSW----- 282
 DB 244 ITDPGKVVTTPLTVGFEDK--DGLQQ-----DARLLKVVSYLDTDGYNMHPENLVA 296
 QY 283 ---LKS-----VPPGPAP-----PLOF-YPGPRFSVQSGSVAS 314
 DB 297 VYDLEAKKILKIEEGVIVPEPRPYDGRDNRAPVAKPLEITEPRGKNTITIGDTIHQ 356
 QY 315 LWTFSGLGAFSGPRIFDYRFQS---ERLVEISLEALAIYGNSPAAMTRYVDCG- 369
 DB 357 NMDPFLRLSRGPIPLSTYNDNGRKQVMDESGMIVPGDDVGMFKAYLIDSGD 416
 QY 370 FGKGYTPTLNGVDCPYATYVDHFLLESQAPKTRDAFCVEQNGGLPLRRHSDLY 429
 DB 417 YGMGTLSPIVGKDAPSNAVLIDETIADYTGKPTIPGVAIFE-----RYAGPEY 468
 QY 430 SHYFGS-----LAETLVYRSMSTLNDYVMDTVPHSGAIEIRFATGYISSAFLFGA 484
 DB 469 KHEMGKPNVSTERRELVAWVWISTVGNVYIFDMVHDNGTIGDAGATIEAVKGLAK 528
 QY 485 T-----GKYNQVSEHTLGTVTHSAHEKVDLDVAGLEWVAEDMVEVPAVPS 535
 DB 529 TMHDPKAKEDTRGLIDINIVGTHQHYNFRDLVDGENTLVAMPVYKPT---A 585
 QY 536 PEHQLORLOVTRKLEMEQAPFLVGSATPRVYLYASNHS--NKMGHPGRIQMLSENG 593
 DB 586 GGPRTSTMOVNOTYIDSEKAAQKPDPTIR---LLSTNSKERNMNPVSY-QIIPYAG 640
 QY 594 --EPLQNSSM-----RGFSWERYQLAVTQKEE-----PSSSVFNQNDWAPT 638
 DB 641 GHHPATGAKFAPDEWIRLSFMDKOLWTRHPTRERPEGYPRNSAHDTGLGYAK- 699
 QY 639 VDFSEFINNETIAGKDLVAVMTAGFLIPHAEDIPRTYVGVNGVGFLLRPNFFEDPS 697
 DB 700 -----DDESLTNHDDVWVITGTTHVARAEEMPMPT--EWALLALKPMNFEDPT 749
 RESULT 10
 AMO_ECOLI STANDARD; PRT; 757 AA.
 ID P46883; P78153; O53008;
 AC 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Copper amine oxidase precursor (EC 1.4.3.6) (Tyramine oxidase)
 DE (2-phenylethylamine oxidase).
 GN TYNA OR MAOA OR B1386.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 NCBI_TaxID=562;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Azakami H., Yamashita M., Roh J.-H., Suzuki H., Kumagai H.,
 RA Murooka Y.;
 RT "Nucleotide sequence of the gene for monamine oxidase (maoA) from
 RT Escherichia coli.";
 RL J. Ferment. Bioeng. 77:315-319(1994).
 RN [2]
 RN SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RC STRAIN=K12 / K10;
 RX MEDLINE=9616434; PubMed=8591028;
 RA Parsons M.R., Convery M.A., Wilmot C.M., Yadar K.D.S., Blakeley V.,
 RA Corner A.S., Phillips S.E.V., McPherson M.J., Knowles P.F.;
 RT "Crystal structure of a quinoxaline: copper amine oxidase of
 RT Escherichia coli at 2-A resolution.";
 RL Structure 3:1171-1184(1995).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;

RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE-97251357; PubMed-9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashiwato K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
 RA Tadamori H., Takeda J., Takekoshi K., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [5]
 RP SEQUENCE OF 1-59 FROM N.A.
 RC STRAIN-W / ATCC 11105;
 RX MEDLINE-98421522; PubMed-9748275;
 RA Fernandez A., Minambres B., Garcia B., Olivera E.R., Luengo J.M.,
 RA Garcia J.L., Diaz E.;
 RT "Catabolism of phenylacetic acid in *Escherichia coli*. Characterization
 RT of a new aerobic hybrid pathway.";
 RL J. Biol. Chem. 273:25974-25986(1998).
 RN [6]
 RP SEQUENCE OF 477-757 FROM N.A.
 RC STRAIN-W / ATCC 11105;
 RX MEDLINE-97263463; PubMed-9109378;
 RA Fernandez A., Prieto M.A., Garcia J.L., Diaz E.;
 RT "Molecular characterization of *Pada*, a phenylacetaldehyde
 RT dehydrogenase from *Escherichia coli*.";
 RL FEBS Lett. 406:23-27(1997).
 RN [7]
 RP SEQUENCE OF 31-50.
 RX MEDLINE-96213037; PubMed-8631685;
 RA Yamashita M., Azakami H., Yokoro N., Roh J.-H., Suzuki H.,
 RA Kumagai H., Murooka Y.;
 RT "maob, a gene that encodes a positive regulator of the monoamine
 RT oxidase gene (*maoa*) in *Escherichia coli*.";
 RL J. Bacteriol. 178:2941-2947(1996).
 RN [8]
 RP SEQUENCE OF 31-40.
 RC STRAIN-K12;
 RX MEDLINE-97195795; PubMed-9043126;
 RA Hanton S.P., Hill T.K., Flavell M.A., Stringfellow J.M., Cooper R.A.;
 RT "2-phenylethylamine catabolism by *Escherichia coli* K-12: gene
 RT organization and expression.";
 RL Microbiology 143:513-518(1997).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE-97200715; PubMed-9048544;
 RA Wilmut C.M., Murray J.M., Alton G., Parsons M.R., Convery M.A.,
 RA Blakeley V., Corner A.S., Palcio M.W., Knowles P.F., McPherson M.J.;
 RT "Catalytic mechanism of the quinonoylamine amine oxidase from
 RT *Escherichia coli*: exploring the reductive half-reaction.";
 RL Biochemistry 36:1608-1620(1997).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE-99315198; PubMed-10387067;
 RA Murray J.M., Sayseil C.G., Wilmut C.M., Tambyrajah W.S., Jaeger J.,
 RA Knowles P.F., Phillips S.E.V., McPherson M.J.;
 RT "The active site base controls cofactor reactivity in *Escherichia*
 RT *coli* amine oxidase: X-ray crystallographic studies with mutational
 RT variants.";
 RL Biochemistry 38:8217-8227(1999).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).

RX MEDLINE-20045000; PubMed-10576737;
 RA Wilmut C.M., Hajdu J., McPherson M.J., Knowles P.F., Phillips S.E.;
 RT "Visualization of dioxygen bound to copper during enzyme catalysis.";
 RL Science 286:1724-1728(1999).
 CC -1- CATALYTIC ACTIVITY: 2-phenylethylamine + H(2)O + O(2) =
 CC phenylacetaldehyde + NH(3) + H(2)O(2).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion, 2 calcium ions and 1 topaguinone per
 CC subunit.
 CC -1- PATHWAY: INITIAL STEPS OF 2-PHENYLETHYLAMINE CATABOLISM.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- PFM: Topaguinone (TPQ) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D23670; BA004900.1; -;
 DR EMBL: L47571; AAC37012.1; -;
 DR EMBL: AE000235; AAC74468.1; -;
 DR EMBL: D90777; BA014966.1; -;
 DR EMBL: D90776; BA014993.1; -;
 DR EMBL: X97452; CA66104.1; -;
 DR EMBL: X97453; CA66107.1; -;
 DR PDB: 1OAC; 03-APR-96.
 DR PDB: 1SPU; 12-MAR-97.
 DR PDB: 1OAF; 23-AUG-99.
 DR PDB: 1OAF; 24-AUG-99.
 DR PDB: 1OAL; 24-AUG-99.
 DR PDB: 1DYU; 29-FEB-00.
 DR PDB: 1D6U; 02-FEB-00.
 DR PDB: 1D6V; 02-FEB-00.
 DR PDB: 1D6Z; 02-FEB-00.
 DR Ecogene: EG13140; tyna.
 DR Interpro: IPR000269; CuNH oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxidn2; 1.
 DR Pfam: PF02728; Cu_amine_oxidn3; 1.
 DR PROSITE: PS01164; COPPER AMINE OXID_1; 1.
 DR PROSITE: PS01165; COPPER AMINE OXID_2; 1.
 KW Oxidoreductase; Metal-binding; Copper; Calcium-binding; TPQ;
 KW Periplasmic; Signal; 3D-structure; Complete proteome.
 FT SIGNAL 1 30
 FT CHAIN 1 30
 FT CA_BIND 31 757 COPPER AMINE OXIDASE.
 FT CA_BIND 563 563 1.
 FT CA_BIND 564 564 1 (VIA CARBOXYL OXYGEN).
 FT CA_BIND 565 565 1.
 FT CA_BIND 708 708 1.
 FT CA_BIND 709 709 1 (VIA CARBOXYL OXYGEN).
 FT CA_BIND 603 603 2.
 FT CA_BIND 700 700 2.
 FT CA_BIND 702 702 2.
 FT CA_BIND 709 709 2 (VIA CARBOXYL OXYGEN).
 FT ACT_SITE 413 413 CATALYTIC BASE (PROBABLE).
 FT MOD_RES 496 496 TOPAGUINONE.
 FT METAL 554 554 COPPER.
 FT METAL 555 555 COPPER.
 FT METAL 719 719 COPPER.
 FT METAL 719 719 COPPER.
 FT VARIANT 42 42 K -> E (IN STRAIN W).
 FT VARIANT 59 59 L -> I (IN STRAIN W).
 FT CONFLICT 33 33 G -> E (IN REF. 7).
 FT CONFLICT 248 248 G -> E (IN REF. 1).
 FT CONFLICT 258 259 GY -> VI (IN REF. 1).
 FT CONFLICT 276 276 I -> II (IN REF. 1).
 FT CONFLICT 288 288 MISSING (IN REF. 1).
 FT CONFLICT 288 288

FT CONFLICT 290 290 P -> I (IN REF. 1).
FT CONFLICT 456 456 A -> P (IN REF. 1).
FT CONFLICT 659 659 H -> D (IN REF. 1).
SQ SEQUENCE 757 AA; 84378 MW; 65600DCE35243DB CRC64;

Query Match 9.9%; Score 397.5; DB 1; Length 757;
Best Local Similarity 27.3%; Pred. No. 2.5e-23;
Matches 142; Conservative 74; Mismatches 211; Indels 89; Gaps 20;

QY 246 FYQGRHYDSLAQLEAFAGLVNVLIPNNGTSGSM-----LKS 285
D 259 YFDGR--DGLNQ-----DARLKVSYLDVGGNVMARIENLVAVNLEOKKIYIEG 311
QY 286 PVPPEP-----AP---PLQFY--PGSPRSVOGSRVASSLMTFSGLAFSGPR 329
D 312 PVVPVPMARPEPDGRDRAVPAVKPMQIIEPEGNKNTITGDMIHMMNMFHLSMNSRVGPM 371
QY 330 IFDVAFQ-----ERLYVEISLQELALAYGNSPAMATRYVNG--FGMCKYTTPLRGVD 384
D 372 ISTVYNDNGTRKRVMTGSLGKMTVPYGDPIGMYFKAYLDSDYGMGTLTSPILAKGD 431
QY 385 CPYLATYVDMHLESOAPKTRDAFCVEQNOGLPLRRHSDLYSHFEG-----LAET 439
D 432 APSNVLNETIADYTGVPMEIPRAIVFE-----RYAGPEKHQMGQPNVSTERR 483
QY 440 VLVAWSMSTLNTDYVMDTVFHPGSAIEIRYATGY-----ISSAFLEGATSK---YGN 490
D 484 ELVVMISTVGNVDYIFPMIFEHENGITIGDAGTIEAVKGAKTMDHETAKDTRYG 543
QY 491 QVSEHTLGVTHSHAFVVDLVAGLENNVMAEDVPMANPMSPEHQLQVOTYRKL 550
D 544 LIDHNVITGTHOITNFRDLVDGDNNSLVAMPVAKPNT---AGPRTSTMOVNOQNI 600
QY 551 EMEQOAFVLSGATPRRYLYLSNHSNKMGHPRGYRIOMLSFGG--EPLPONSMA----- 603
D 601 GNEQOAAQKFPDGTIR--LLSNKKNRBMGNPVSY--QIIPVAGHHPVAKGAQAPDEMT 657
QY 604 -RGFEMERYQLAVTORKEEPESSSVF--NONDPAPYVDFSPFINNETIAGDLYAVWYA 661
D 658 YHRLSEFMKQLWVTRYHPEERPEEGKYPNRSYHDTGLQGYSK--DNESLNTDAVVMWT 715
QY 662 GELHPIHAEDIPNTVTVGNGVGFLLRPYFDEDEPS 697
D 716 GTTHVARAEEMPIPT--EMVHTLLKPMFDEPT 749

RESULT 11
AMOH_ARTGO STANDARD: PRT: 683 AA.
AC 059118:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histamine oxidase (EC 1.4.3.6) (Copper amine oxidase).
OS Arthrobacter globiformis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=1665;
RN
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN-IFO 12137 / ATCC 8010;
RX MEDLINE=95181469; PubMed=7876243;
RA Chou Y.-H., Matsuzaki R., Fukui T., Shimizu E., Yorifuji T., Sato H.,
Ozaki Y., Tanizawa K.;
"Copper/topa quinone-containing histamine oxidase from Arthrobacter
globiformis. Molecular cloning and sequencing, overproduction of
precursor enzyme, and generation of topa quinone cofactor.";
RL J. Biol. Chem. 270:4712-4720(1995).
CC -1- FUNCTION: OXIDIZE HISTAMINE. OTHER AMINES INCLUDING
PHENETHYLAMINE, TYRAMINE, TRYPTAMINE, PUTRESCINE, AND BENZYLAMINE
ALSO SERVE AS SUBSTRATE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
H(2)O(2).

CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: BY HISTAMINE.
CC -1- PM: Topaquinone (TPQ) is generated by copper-dependent
autooxidation of a specific tyrosyl residue.
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPOQUINONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC
DR EMBL: D38508; BAA07517.1; -;
DR HSSP: P46881; 1A7A.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; FALSE_NEG.
KW Oxidoreductase; Copper; TPQ; Metal-binding.
FT INIT_MET 0
FT MOD_RES 401 401 TOPOQUINONE.
FT METAL 450 450 COPPER (POTENTIAL).
FT METAL 452 452 COPPER (POTENTIAL).
FT METAL 611 611 COPPER (POTENTIAL).
SQ SEQUENCE 683 AA; 74979 MW; 233C46C017201AAO CRC64;

Query Match 9.4%; Score 377; DB 1; Length 683;
Best Local Similarity 24.1%; Pred. No. 8.8e-22;
Matches 182; Conservative 107; Mismatches 343; Indels 124; Gaps 28;

QY 20 PHCPSVSPSAQWPT--HPOQSOLFADLSREELTAVMRELTQSLGSLVAAQAPSDNCV 77
D 6 PSTPLVQDPVPVATLVHAAQHPLEOLSAEETHEARRILAE--AGLV----- 51
QY 78 FSEVQLPPEKAALALHLDGRSP-----PAREALATVFFGRPO-----PNVSE 121
D 52 -----ESTRPAVLGLIEPPKTRTQGDVGAALVAVMLDAQNSLDVRLSLANG 102
QY 122 LVVGPLPHPSYKRDVTEHNGGPLPYHRRPVYFOETLDDQMTFNRELFOASGLLHCCF 181
D 103 LVVD-----RRELNPEDG-----QLPVLEEFGIILEDL--SEDPOMNAL----- 142
QY 182 YKHGRNLVTMTAP-----RGLQSGDRATWFGLYNISGAFPLHVGLELVNHA 234
D 143 -TARGILTPAQVAVAPLSAGVFEYGNNEGKRL-----LRGLGFQDHP-----ADHPV 188
QY 235 LDPARWITQKVEFYQGRHYDSLAQLEAFAGLVNVLIPD--NGTGSWSLSKSPVGPAP 293
D 189 AHPIDGLVAFVDEVNRVNH-----IDGVPV---PEVGNVYDPAIRGELRDLPL 239
QY 294 PLQFYPOGRRESVOGSRVASSLMTFSFGIGAFSGPIFDYR--FOGER--LYEISLQEA 349
D 240 IEMQPEGSPFLTEGNHLSWAGMDLVGFDAREGLVHQLHSHKRRRPVHIRASISEM 299
QY 350 LAIVGNSPAMATTRYVDG--FGMGKYYTPLRGVYCPPLATYVDMHFLLEQAPRTID 408
D 300 VVPYGDPSYRSWMONTFDSGETLVGRDANSRLGCDLDTIYMSVVAADGPNPTTN 359
QY 409 AFCEVEONGLPLRRHSDLYSHFGLAETVLVVAWSMSTLNTDYVMDTVFHPGSAIEI 468
D 360 GICHEEDAGLWIK--HTDEWAGSEVRRNRRLRVYSFTVGNVDGFWYLYLDQETIF 417
QY 469 RRYATGYSIAFLFGATGTYGNQVSEHTIGYHTHSANHKYDLDVAGLENNVMAEDMVY 528
D 418 EAKAGIYETTALPKDYAVASEIAPGLGAPYHOHLFSARLDMYIDGIANRVEEDLVRL 477
QY 529 PAAVWMSPEHQLQVOTYRKLLEMEQOAFVLSGATPRRYLYLSNHS--KKMGHPRGYRQ 587

Db 478 PKG-PGNHG--NAFTQRTLLARESEAVRDADGAKGRVWHISNPDLSLHGHGPGYTLY 534
QY 588 MLRAGEP---LPONSSMARGFSEWRYOLAVYORKEEPPSSSVFNQNDPMAPTVDESD 644
Db 535 P---EGNPTLLMADDSSTASRAFAARHMLWYTRHAEELLYAGDVPNOHPCGAVLP-AYV 590
QY 645 INNETHACKDLVAVYTAGFLIHPHAEIDPNTVYVNGVGFLLRPVNFEDDESPFSADSI 704
Db 591 AODDIDGQDLVYVHSHFGLTHFPREDMP---IMPVDTGTFLKPGFDEPNTLVNPSSA 648
QY 705 YFRGDODAGACEVNPPLACLPQAAACAPDLPAFESHOG 740
Db 649 -----AGHCGTG-----SEREHAAPGCTAVGHSG 672

RESULT 12
AMOI_ARTSI STANDARD; PRT: 648 AA.
ID AMOI_ARTSI
AC 007121;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase precursor (EC 1.4.3.6) (MAOXI).
GN MAOI.
OS Arthrobacter sp. (strain P1).
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=47915;
RN [1]
RP MEDLINE=93374858; PubMed=8366046;
RA Zhang X., Fuller J.H., McIntire W.S.;
RT "Cloning, sequencing, expression, and regulation of the structural gene for the copper/topa quinone-containing methylamine oxidase from Arthrobacter strain P1, a Gram-positive facultative methylotroph.";
RL J. Bacteriol. 175:5617-5627(1993).
CC -1- FUNCTION: THE EXACT FUNCTION OF MAOXI IS NOT KNOWN.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) + H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC -1- PFM: Topaquinone (TPQ) is generated by copper-dependent autooxidation of a specific tyrosyl residue (By similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPOQUINONE OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC
CC EMBL: L12983; AAA22076.1; .
DR HSSP: P46881; IAV4.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxid2_1.
DR Pfam: PF02728; Cu_amine_oxid3_1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
KW Oxidoreductase; Copper; TPQ; Metal-binding.
FT PROPEP 1 9
FT CHAIN 10 648 COPPER AMINE OXIDASE
FT MOD_RES 385 385 TOPAQUINONE (BY SIMILARITY).
FT METAL 436 436 COPPER (POTENTIAL).
FT METAL 438 438 COPPER (POTENTIAL).
FT METAL 595 595 COPPER (POTENTIAL).
SQ SEQUENCE 648 AA; 72760 MM; B2F9E267492253B5 CRC64;

Query Match 9.18; Score 363; DB 1; Length 648;
Best Local Similarity 24.08; Pred. No. 1e-20;
Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

QY 36 GQSQFLADLSREELAVMRFILQRLGRLVDAAQARSDNCVFSEYLOLPPKAAALAHLD 95
Db 12 GVSHPLDLSREELARAVALIKE--GPAAESFR-----FISLELRPSKDDL----- 57
QY 96 RGSPPAREALAVYFFGQPOPNWSELVYGPLPHPSYRDVYVERHGRLPYHRRPVLFQ 155
Db 58 RAGVAVAEADAVLV--DRAQRSEAVVD-----LEAGTVDSKLLAENIQPFPLMD 108
QY 156 EYLDIDQIFNRELIPQASGLLHCCFYKRGKRNLTMTTARGLQSGDRA---TWFGLYY 212
Db 109 EFACEDAC--RKDE-----VIALAKRGLNLNLDVCFEPPSVGYE 148
QY 213 NISGAGFLVHVGLELVNHNKALD-PARKTIQK--VFYQGRYDSLAOLEAFEGAVLVY 269
Db 149 GEDNEGRL--WRALVFVRDADSPYAHPIENIFYEY-----DLNAG--KV 191
QY 270 VLIPIVNGTGSMSLSKSPPPGAPAPLQFY-----POGRFSVQSGRVASS 314
Db 192 VRLIEDQA-----IPPSARGNTLPYVGEARLDKLPNLITQDEGSAFYTGHNHYWA 244
QY 315 LWTFSFGLGAFSGPRIFDVRQGE---RLYEISLQALAIYGSNSPA-AMTTRVYDGG 369
Db 245 DMSFVGFPTREGVLVHOLKFKDQGVDRPVINRASLSMWVVPYGDVAVQAKKNAFDSGE 304
QY 370 FGMKRYTPTLRGVDCPYLATYVDMHFLLESQAPKTIADACVFEQONGLPRLRHSDLY 429
Db 305 YNIGMANSLLTGCCDCELEIRYFDGHSVDSHGNPTINAIICMHEEDSI-----LW 356
QY 430 SH--YFGGLAET---VLVRSMSSTLVNDYVMDVFPSPGAIEIRFATGYISSA-FLF 482
Db 357 KHFDRREGTAETERRKRLVISTATVAVYEAFAFYHNLDSIELVATGILSTAGLPL 416
QY 483 GATGRYGNQVSEHTL-GTVTHSAHFYVDLVAGLENNVMAEDVYFVMAVPMWSEHQL- 540
Db 417 GEKNYGGSLNNDGGLYAAIHQHFVVRMDFELDGVKNAVYEVDMEY-----PEHNPT 468
QY 541 -QRQVATKLLMEERQAFVIGSAPRYLYLASNS--KKMGHPRKRYRQMLSFAGEPLPQ 598
Db 469 GTAFNAVDRLLLETBOKAIAKRTNEAKHFRWKIANHSHKLLVNEPAVRL-----IPT 519
QY 599 NSSMARGFSEWRYOLAVYORKEEPPSSSVFNQNDPMAPTVD-----FSDFINNET--- 649
Db 520 NG-----IQDLA--RDDAVYSKRAQFARNNLWTAVDRTERFAAAGEPNOATGAD 567
QY 650 -----IACKDLVAVYTAGFLIHPHAEIDPNTVYVNGVGFLLRPVNFEDDEPS 697
Db 568 DGLHMTQKDRNIYDTDLVWYVTFGMHVVRLDMP--VPMRONIGFMLEPHGFENQNP 625
QY 698 FYSADSIYFRGDQDAGAC 715
Db 626 LNLPTSTTGTGEADTC 643

RESULT 13
AMOI_ARTSI STANDARD; PRT: 648 AA.
ID AMOI_ARTSI
AC 007123;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper methylamine oxidase precursor (EC 1.4.3.6) (MAOXII).
GN MAOI.
OS Arthrobacter sp. (strain P1).
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=47915;
RN [1]
RP MEDLINE=93374858; PubMed=8366046;
RA Zhang X., Fuller J.H., McIntire W.S.;
RT "Cloning, sequencing, expression, and regulation of the structural gene for the copper/topa quinone-containing methylamine oxidase from Arthrobacter strain P1, a Gram-positive facultative methylotroph.";
RT Arthrobacter strain P1, a Gram-positive facultative methylotroph.";

RL J. Bacteriol. 175:5617-5627(1993).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) + H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- INDUCTION: BY METHYLAMINE.
 CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent autooxidation of a specific tyrosyl residue (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUNINE OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L13990; AAA22074.1; -
 DR HSP, P46881; 1A4.
 DR InterPro: IPR000269; CUNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.
 DR Pfam: PF02728; Cu_amine_oxid3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 KW Oxidoreductase; Copper; TPO; Metal-binding.
 FT PROPEP 1 9
 FT CHAIN 10 648 COPPER METHYLAMINE OXIDASE.
 FT MOD_RES 385 385 TOPAQUNINE (BY SIMILARITY).
 FT METAL 436 436 COPPER (POTENTIAL).
 FT METAL 438 438 COPPER (POTENTIAL).
 FT METAL 595 595 COPPER (POTENTIAL).
 SQ SEQUENCE 648 AA; 72805 MM; R2FB2787492253B5 CRC64;
 Query Match 9.18; Score 363; DB 1; Length 648;
 Best local similarity 24.08; Pred. No. 1e-20;
 Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;
 36 GOSQFLADLSRELTAVNRFLTQRLGVLDAQAQRPSCNCFVSEVLQPPRAALAHLD 95
 12 GVSHPDLDSVEIARAVALIKE--GPAASESFR-----FISVLEEPSKDDL----- 57
 96 RGSPPAREALAIYFFGQOPNSELVVGRLPHPSYMRDLYVEHNGGLPRHRRPVLEQ 155
 58 RAGVAVAEADAVLV--DRAQARSEAVD-----LEAGVDSWKLLAENIQPPFMD 108
 156 EYLDIDQMFNRELPOASGLLHHCFFYKRGHNLVTMTTAPGLQSGDRA--TWFGLYY 212
 109 EFAEEDAC--RKQPE-----VIALAKGLNLDLVCPEPSVGF 148
 213 NISGAGFLHVGLELLVNHKALD--PARMTQK--VFYQGRYDLSLAOLEAFEGALVNV 269
 149 GEDNGGRRL--MRALVYFDEADSPYAHPIENFIVFY-----DLNAG--KV 191
 270 VLIPIINGGGSLSKSPYPPGAPPLQFY-----PQGRFVSQGSRVASS 314
 192 VRLIEDQA-----IPVPSARGNLYPYVGEARTDLPLNTQDEGASFTYTGHNVTWA 244
 315 LMTFFGGLGAFSGPRIFDFRQGE-----RLVYEISLQELATLYGGSNPA--AMTRVYDVG 369
 245 DMSFVGTGTPREGVLVHQLKFDQGVDRVYNRASLSENVYVYGTAPVQAQAKNAFDSGE 304
 370 FGMGYTTPPLRGVDCPYLATVDMHFLLESQAPRTINDAFVFEQONGLPLRRHSDLY 429
 305 YNIGMANSILTGCDLCEIKFYDGHSDVSHGNPMTIEANICMHEEDSI-----LM 356
 430 SH--YFGGLAET-----VLVYRSMSTLNTDYMDVFNHSGAIEIRFYATGYSN-FLF 482
 357 KHFDREGAETRRSRKLVLSIFATVANYEVAFYVHMLFDGSIIEFLVKATGILSTAGQLP 416
 483 GATGKGNQVSEHTL-GVHTSHAHFKVDLDVAGLENNWADNMFVPAVWSPENHQL- 540

DB 417 GERNPYGOSLNNDGLIAPRHOHMFVNRDFFELDGYKNAVVEVDMEY-----PENHPT 468
 QY 541 -QRLOVTRKLLMEEQAAFLVGSATPRYLYLASNHS-NKMGHPRCYRIQMSFAGEPPLQ 598
 DB 469 GTAFMAVDRLLETQKAIKRTKNEAKHRPMKIANHESKMLNEVPVYRL-----IPT 519
 QY 599 NSSARGSWERYQLAVYQRKEEPPSSSVYRNDPAPTYD-----FSDFINNET--- 649
 DB 520 NG-----IDLAA--RODAYVSKRAQFARNLWLTAYDRTERFAAGEYPNQATAD 567
 QY 650 -----IAGKDLVAVWTAGFLIHPAEDIPNTVYVGNVGFLLRPYFDEDDPS 697
 DB 568 DGLHNTQKDNRYDITDLVWYTFEGMHVYRLDMP--VMPRONIGFLEPHGFENQNP 625
 QY 698 FYSADSIYFRDQDAGAC 715
 DB 626 INLPSTSTQOTGEADTC 643
 RESULT 14
 ID PAOX_ARTGO STANDARD; PRT; 638 AA.
 AC P46881;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phenylethylamine oxidase precursor (EC 1.4.3.6) (amine oxidase).
 OS Arthrobacter globiformis.
 CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 CC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
 OX NCBI_TaxID=165;
 [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=IFO 12137 / ATCC 8010;
 RX MEDLINE=94197690; PubMed=8147851;
 RA Tanizawa K., Matsuzaki R., Shimizu E., Yorifuji T., Fukui T.;
 RT "Cloning and sequencing of phenylethylamine oxidase from Arthrobacter globiformis and implication of Tyr-382 as the precursor to its covalently bound quinine cofactor."
 RT Biochem. Biophys. Res. Commun. 199:1096-1102(1994).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 9-628.
 RX MEDLINE=98070236; PubMed=9405045;
 RA Milne M.C., Dooley D.M., Freeman H.C., Guss J.M., Matsumami H.,
 RA Wallace M.S., Ruggiero C.E., Tanizawa K., Yamaguchi H.;
 RT "Crystal structures of the copper-containing amine oxidase from Arthrobacter globiformis in the holo and apo forms: implications for the biogenesis of topaquinine."
 RT Biochemistry 36:16116-16133(1997).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) + H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- INDUCTION: BY PHENETHYLAMINE.
 CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent autooxidation of a specific tyrosyl residue.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUNINE OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U03517; AAA18114.1; -
 DR PDB: 1AVK; 18-MAR-98.
 DR PDB: 1AVL; 25-MAR-98.
 DR InterPro: IPR000269; CUNH_oxidase.
 DR Pfam: PF02727; Cu_amine_oxid; 1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.

DR Pfam: PF02728; Cu:amine_oxidn3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 KM Oxidoreductase; Copper; TPQ; 3D-structure; Metal-binding.
 FT PROPE 1 2
 FT CHAIN 3 638 PHENYLETHYLAMINE OXIDASE.
 FT MOD_RES 382 382 TOPAQUINONE.
 FT METAL 431 431 COPPER.
 FT METAL 433 433 COPPER.
 FT METAL 592 592 COPPER.
 FT DISULFID 317 343
 FT MUTAGEN 382 382
 SQ SEQUENCE 638 AA; 70646 MW; 1800396BA7A983F2 CRC64;

Query Match 8.5%; Score 340; DB 1; Length 638;
 Best Local Similarity 23.7%; Pred. No. 6,3e-19;

Matches 157; Conservative 100; Mismatches 274; Indels 132; Gaps 28;

DB 87 KAALAHLD--RGSPPPREALAIVF---GROPQPNVSELVGFLPHPSYMRDVTVER 140
 38 RLAIVGLVDPAKGASEADRRFRVFIHVSARQ-----EVTVSV 79
 141 HGGPL-----PYHRRPVLFQEYLDIDQMFN-----RELQAS 173
 80 TNGTVISAVELDTAATGELPYLEEFVEYEQLLATDERWLKALAARNLDVSKVRAPLSA 139
 174 GLHHCCEYKRGRLVMTAPRGL---QSGDRATWGLYNTSGAGFLHNGLELL 229
 140 GVFEFA---EEGRKIL-----KGLATVQDFPEDSAM---AHPVDGLVAVDVVSKETV 187
 230 VNHKALDPAKMTIQVYOGRYDLSLAOLEAFEAGLVNVVLI PDNGTSGSLSKSPVP 289
 188 ---RVIDGVFVPA--EHGNTDP-----ELGPLMT 215
 220 GAPRLOVYPOGPRRSVQ--GSNVASSLMTFSGLGAFSGPRFVDVRFQ--GERL---VYEI 344
 216 TOKPISITQPEGSPFTVGGNHIEKEKSLDVGFVREGVVLHNTAFRGDRLRPIINRA 275
 345 SLOEALATIGGNSPAAMTTRYVD--GFGMGKYYTTPLTRGVDC---PYATVVDHMFLE 399
 276 SLAEVNVYVYGDSPIRSNQNTDTEGEYLVGVANSLELGCDLGDITVLSPIVSDAF--- 332
 400 SOAKPTIDAFCEVRONOGLPLRRHNSDLSHYFGGLAETVLVNSMSLTLNVYDVMTV 459
 333 -GNPREINNGICMHEEDGILAK--HSDLMGINTYRRRRRVVISPFTIGNVYDGFVY 389
 460 FHPGSAIEIRFATGYI--SSAFLGATGKYGNQVSEHNLGTVHTSHAKKVLVDVAGLEN 518
 330 LLDGTIEPEAKATGVFTSAFPEGSDNI--SOLAPGLGAPPHOHIFSAFLDMAIDGFTN 448
 519 WVMADNVFVPAVPMSEHQLRLOVTRKLEMEEOAFLVGSATPRVLYLASNHS--NK 577
 449 RVEEDVYRQTGPNENGNAFSR---KRTVLTRESEAVREADARTGRWITSNPSKNR 505
 578 WCHPRGYRI---QMLSPAGEPLPNSSMARGFSWERYOLAVTQRKEEPPSSSVF--NQND 633
 506 LNEPVGYRLAHNQPTLLADP---GSSIAARRAFAKTKDLMTVRYADDERYPTGDFVNGHS 562
 634 PMAPTVDSDFINNTTIGKDLVAVTAGFLHIAEDIPNTVTVGNGVGEFLRPRNFED 693
 563 GGAGLPSR--IAQDRDIDGQDVLVWHTGELTHFPRVEDMP--IMPVDVYGFKLREGEFD 618
 694 EDP 696
 619 RSP 621

RESULT 15

AMO_PEA STANDARD; PRT; 674 AA.

ID AMO_PEA
 AC 043077;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DR 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Amine oxidase [copper-containing] precursor (EC 1.4.3.6).
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 OX NCBI_Taxid:3888;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Tipping A.J., McPherson M.J.;
 RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
 RN (2)
 RC X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE-96398683; PubMed=8805580;
 RA Kumar V., Doolley D.M., Freeman H.C., Guss J.M., Harvey I.,
 RA McGuffil M.A., Milce M.C., Zubak V.M.;
 RT "Crystal structure of a eukaryotic (pea seedling) copper-containing
 RT amine oxidase at 2.2-A resolution."
 RL Structure 4:943-955(1996).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion, 1 manganese ion and 1 topaquinone
 CC per subunit.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- PFM: Topaquinone (TPQ) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L39931; AAA62490.1; -
 DR PDB: 1KST; 24-DEC-97.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu:amine_oxid; 1.
 DR Pfam: PF02727; Cu:amine_oxidn3; 1.
 DR Pfam: PF02728; Cu:amine_oxidn3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; FALSE_NEG.
 KM Oxidoreductase; Metal-binding; Copper; Manganese; TPQ; glycoprotein;
 KM Signal; 3D-structure.
 FT SIGNAL 1 25
 FT CHAIN 26 674 AMINE OXIDASE [COPPER-CONTAINING].
 FT METAL 476 476 MANGANESE.
 FT METAL 477 477 MANGANESE (VIA CARBOXYL OXYGEN).
 FT METAL 478 478 MANGANESE.
 FT METAL 617 617 MANGANESE.
 FT METAL 618 618 MANGANESE (VIA CARBOXYL OXYGEN).
 FT MOD_RES 412 412 TOPAQUINONE.
 FT METAL 467 467 COPPER.
 FT METAL 469 469 COPPER.
 FT METAL 628 628 COPPER.
 FT DISULFID 162 183
 FT DISULFID 344 370
 FT CARBOHYD 156 156 N-LINKED (GLCNAC...) (PROBABLE).
 FT CARBOHYD 389 389 N-LINKED (GLCNAC...) (PROBABLE).
 FT CARBOHYD 583 583 N-LINKED (GLCNAC...) (PROBABLE).
 SQ SEQUENCE 674 AA; 76358 MW; 30735390071DD18E CRC64;

Query Match 8.0%; Score 320; DB 1; Length 674;

Best Local Similarity 25.8%; Pred. No. 2,6e-17;

Matches 120; Conservative 66; Mismatches 189; Indels 90; Gaps 20;

DB 282 SLKSPVPGPAP--LQYPOGPRRSVQSGKVASLMTFSGLGAFSG-----PRIFVY- 333
 233 SKQSP-PFGPKOHSILTSHPQPGFQINGHSVAMNKRFGIFGVRAIVISLASYDLE 291

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 18.4058 Seconds
(without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179
Sequence: 1 MSPILGKWKIKGLVOPTRL.....IAMPLOGMATFGGDHPK 218

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq.101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1179	100.0	218	AAW84205	Protein comprising
2	1179	100.0	218	ABB09809	Glutathione-S-tran
3	1179	100.0	224	AAW02122	Glutathione-S-tran
4	1179	100.0	225	AAW02123	Glutathione-S-tran
5	1179	100.0	232	AAW72793	Glutathione-S-tran
6	1179	100.0	240	AAW73909	GST protein sequen
7	1179	100.0	245	AAW31612	Amino acid sequenc
8	1179	100.0	247	AAW74114	GST/GI transport r
9	1179	100.0	247	AAW74120	GST/GI transport r
10	1179	100.0	247	AAW74103	GST/GI transport r

11	1179	100.0	248	AAW74112	GST/GI transport r
12	1179	100.0	248	AAW74113	GST/GI transport r
13	1179	100.0	248	AAW74121	GST/GI transport r
14	1179	100.0	248	AAW74122	GST/GI transport r
15	1179	100.0	248	AAW74100	GST/GI transport r
16	1179	100.0	248	AAW74101	GST/GI transport r
17	1179	100.0	248	AAW74102	GST/GI transport r
18	1179	100.0	249	AAW74115	GST/GI transport r
19	1179	100.0	257	AAW74111	GST/GI transport r
20	1179	100.0	257	AAW74119	GST/GI transport r
21	1179	100.0	257	AAW74124	GST/GI transport r
22	1179	100.0	257	AAW74126	GST/GI transport r
23	1179	100.0	257	AAW74105	GST/GI transport r
24	1179	100.0	257	AAW74109	GST/GI transport r
25	1179	100.0	258	AAW74117	GST/GI transport r
26	1179	100.0	258	AAW74104	GST/GI transport r
27	1179	100.0	259	AAW74130	GST/GI transport r
28	1179	100.0	259	AAW74118	GST/GI transport r
29	1179	100.0	259	AAW74125	GST/GI transport r
30	1179	100.0	259	AAW74110	GST/GI transport r
31	1179	100.0	262	AAW74128	GST/GI transport r
32	1179	100.0	262	AAW74129	GST/GI transport r
33	1179	100.0	267	AAW74106	GST/GI transport r
34	1179	100.0	271	AAW74103	GST/GI transport r
35	1179	100.0	272	AAW56327	Cap-A-B fusion pro
36	1179	100.0	272	AAW74116	Amino acid sequenc
37	1179	100.0	277	AAW74098	GST/GI transport r
38	1179	100.0	277	AAW74099	GST/GI transport r
39	1179	100.0	277	AAW74107	GST/GI transport r
40	1179	100.0	279	AAW74108	GST/GI transport r
41	1179	100.0	279	AAW74097	GST/GI transport r
42	1179	100.0	282	AAW56326	Amino acid sequenc
43	1179	100.0	282	AAW74127	GST/GI transport r
44	1179	100.0	282	AAW74123	GST/GI transport r
45	1179	100.0	282	AAW74095	GST/GI transport r

ALIGNMENTS

RESULT 1	
ID	AAW84205
AC	AAW84205 standard; Protein; 218 AA.
XX	
AC	AAW84205;
XX	
DT	25-MAR-1999 (first entry)
XX	
DE	Protein comprising a purification tag.
XX	
KW	Proteinase site; bone morphogenetic fusion protein; bone binding site;
KW	bone morphogenetic protein; transforming growth factor beta;
KW	active fragment; wound healing; bone growth; purification tag.
XX	
OS	Unidentified.
XX	
PN	W09855137-A1.
XX	
PD	10-DEC-1998.
XX	
PF	02-JUN-1998; 98MO-US11189.
XX	
PR	03-JUN-1997; 97US-0868452.
XX	
PA	(HALL/) HALL F. L.
PA	(HAMB/) HAN B.
PA	(NIMK/) NIMMI M E.
PA	(SHOR/) SHORS E C.
PA	(WUL/) WU L.
PI	Hall FL, Han B, Nimmi ME, Shors EC, Wu L;
XX	
DR	WPI: 1999-059875/05.

DR N-PSDB; AAV99373.
 XX
 PT New bone morphogenetic fusion proteins - comprising a purification
 tag and a bone morphogenetic active fragment, used for enhancing
 wound healing or bone growth
 PT
 XX
 PS Disclosure; Page 39-40; 64pp; English.
 CC The present sequence represents a protein comprising a purification tag
 CC that was used in the creation of the bone morphogenetic fusion proteins
 CC of the invention. The bone morphogenetic fusion protein may contain some
 CC or all of the following elements: a purification tag, a proteinase site,
 CC an ECM/bone binding site, a second proteinase site, and a bone
 CC morphogenetic protein active fragment. The fusion proteins of the
 CC invention also includes proteins that have transforming growth factor
 CC beta active fragments instead of bone morphogenetic protein active
 CC fragments. The bone morphogenetic fusion proteins can be used for
 CC enhancing wound healing or bone growth.
 CC
 SQ Sequence 218 AA;
 Query Match 100.0%; Score 1179; DB 20; Length 218;
 Best Local Similarity 100.0%; Pred. No. 4e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKMRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKMRNKKFELGLEFPNLPYYID 60
 QY 61 GDVLTQSMATIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKV 120
 DB 61 GDVLTQSMATIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKV 120
 QY 121 DFLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 QY 121 DFLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 QY 181 KRIFAIPQIDKYLKSSKIYIAMPLOGMATFGGSDHPK 218
 DB 181 KRIFAIPQIDKYLKSSKIYIAMPLOGMATFGGSDHPK 218
 RESULT 2
 ID ABB09809 standard; Protein: 218 AA.
 AC ABB09809;
 XX
 DT 22-JUL-2002 (first entry)
 DE Glutathione-S-transferase (GST) of Schistosoma japonicum.
 KW Glutathione-S-transferase; GST; enzyme; metallothionein; MT-IIA;
 KM heavy metal contact; UV light.
 XX
 OS Schistosoma japonicum.
 XX
 PN FR2813529-A1.
 XX
 PD 08-MAR-2002.
 XX
 PF 31-JUL-2001; 2001FR-0010239.
 XX
 PR 06-SEP-2000; 2000ES-0000217.
 XX
 PA (PROV-) PROVITAL SA.
 XX
 PI Armengol SR, Benalgués MA;
 XX
 DR WPI: 2002-332437/37.
 PT Topical compositions, used to protect against heavy metals and
 ultraviolet radiation, contain metallothionein proteins -

XX
 PS Example; Page 28-29; 33pp; French.
 XX
 CC The present sequence represents the glutathione-S-transferase (GST)
 CC protein of Schistosoma japonicum. A human metallothionein class IIA
 CC (MT-IIA) was expressed fused to the present GST protein, for
 CC purification purposes. The MT-IIA protein was used to produce
 CC topical pharmaceutical and cosmetic compositions, in the form of gels,
 CC creams, ointments, soaps and body lotions. In the compositions, the
 CC metallothionein proteins have a prosthetic group attached to a zinc
 CC ion. The topical compositions can be used for protection of human or
 CC animal external tissue against the effects of heavy metal contact and
 UV light.
 CC
 SQ Sequence 218 AA;
 Query Match 100.0%; Score 1179; DB 23; Length 218;
 Best Local Similarity 100.0%; Pred. No. 4e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKMRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKMRNKKFELGLEFPNLPYYID 60
 QY 61 GDVLTQSMATIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKV 120
 DB 61 GDVLTQSMATIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKV 120
 QY 121 DFLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 QY 121 DFLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 QY 181 KRIFAIPQIDKYLKSSKIYIAMPLOGMATFGGSDHPK 218
 DB 181 KRIFAIPQIDKYLKSSKIYIAMPLOGMATFGGSDHPK 218
 RESULT 3
 ID AAM02122 standard; Protein: 224 AA.
 AC AAM02122;
 XX
 DT 28-OCT-1996 (first entry)
 DE Glutathione-s-transferase epitope.
 KW Bone resorption disease; osteoporosis; src SH2 domain antagonist;
 KM src homology 2 domain; glutathione-s-transferase; Tyr SH2;
 KW Grb2 SH2; SH-PTP2 SH2.
 XX
 OS Not specified.
 XX
 PN EP727211-A1.
 XX
 PD 21-AUG-1996.
 XX
 PF 07-FEB-1996; 96EP-0200270.
 XX
 PR 29-DEC-1995; 95US-0580868.
 PR 10-FEB-1995; 95US-0386381.
 PR 07-MAR-1995; 95US-0400220.
 PR 30-JUN-1995; 95US-0497357.
 PR 11-OCT-1995; 95US-0541080.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Dunnington DJ;
 XX
 DR WPI: 1996-372674/38.
 PT Use of selective src SH2 domain ligand - to prepare medicament for
 treating bone resorption disease

XX Example 11: Page 31-32; 47pp; English.

PS A glutathione-s-transferase epitope (AAW02122) was used to construct
 CC fusion proteins with human SH2 domains from fyn (see also AAW02124),
 CC Grb2 (AAW02127) and SH-PP2 (AAW02125); a similar epitope (AAW02123) was
 CC used to construct a fusion protein with human p85 SH2 domain (see
 CC also AAW02126). These fusion proteins, and tagged proteins including
 CC other SH2 domains (see also AAW02119-21), can be used in binding
 CC assays to determine the specificity of cpds. to inhibit SH2 domains;
 CC cpds. that selectively inhibit the human src SH2 domain are useful
 CC in treating bone resorption diseases such as osteoporosis.

XX Sequence 224 AA:

Query Match 100.0%; Score 1179; DB 17; Length 224;
 Best Local Similarity 100.0%; Pred. No. 4,1e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPTLGWTKIGLVOPRLLEYLEEYERDEGDMRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPRLLEYLEEYERDEGDMRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTOSMAIIRYIADKHNMLGSCPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLVY 120
 DB 61 GDVKLTOSMAIIRYIADKHNMLGSCPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLVY 120
 QY 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDVLYDALDVLVYMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDVLYDALDVLVYMDPCLDAFPKLVCFK 180
 QY 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDVLYDALDVLVYMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDVLYDALDVLVYMDPCLDAFPKLVCFK 180
 QY 181 KRTEAIPOIDIKYKSSKYIAMPLOGQATFGGSDHPK 218
 DB 181 KRTEAIPOIDIKYKSSKYIAMPLOGQATFGGSDHPK 218

RESULT 4
 AAW02123
 ID AAW02123 standard; Protein; 225 AA.

AC AAW02123;
 XX 28-OCT-1996 (first entry)
 DT
 DE Glutathione-s-transferase epitope.
 XX
 KW Bone resorption disease; osteoporosis; src SH2 domain antagonist;
 KM src homology 2 domain; glutathione-s-transferase; p85 SH2.
 XX Not specified.
 OS
 XX
 PN EF727211-A1.
 PD
 XX 21-AUG-1996.
 PF
 XX 07-FEB-1996; 96EP-0200270.
 PR
 XX 29-DEC-1995; 95US-0580868.
 PR 10-FEB-1995; 95US-0386381.
 PR 07-MAR-1995; 95US-0400220.
 PR 30-JUN-1995; 95US-0497357.
 PR 11-OCT-1995; 95US-0541080.
 XX
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 XX
 PI Dunnington DJ;
 XX
 DR WPI; 1996-372674/38.
 XX
 PT Use of selective src SH2 domain ligand - to prepare medicament for
 XX treating bone resorption disease
 PS Example 11: Page 31-32; 47pp; English.

XX A glutathione-s-transferase epitope (AAW02122) was used to construct
 CC fusion proteins with human SH2 domains from fyn (see also AAW02124),
 CC Grb2 (AAW02127) and SH-PP2 (AAW02125); a similar epitope (AAW02123) was
 CC used to construct a fusion protein with human p85 SH2 domain (see
 CC also AAW02126). These fusion proteins, and tagged proteins including
 CC other SH2 domains (see also AAW02119-21), can be used in binding
 CC assays to determine the specificity of cpds. to inhibit SH2 domains;
 CC cpds. that selectively inhibit the human src SH2 domain are useful
 CC in treating bone resorption diseases such as osteoporosis.

XX Sequence 225 AA:

Query Match 100.0%; Score 1179; DB 17; Length 225;
 Best Local Similarity 100.0%; Pred. No. 4,2e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPTLGWTKIGLVOPRLLEYLEEYERDEGDMRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPRLLEYLEEYERDEGDMRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTOSMAIIRYIADKHNMLGSCPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLVY 120
 DB 61 GDVKLTOSMAIIRYIADKHNMLGSCPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLVY 120
 QY 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDVLYDALDVLVYMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDVLYDALDVLVYMDPCLDAFPKLVCFK 180
 QY 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDVLYDALDVLVYMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDVLYDALDVLVYMDPCLDAFPKLVCFK 180
 QY 181 KRTEAIPOIDIKYKSSKYIAMPLOGQATFGGSDHPK 218
 DB 181 KRTEAIPOIDIKYKSSKYIAMPLOGQATFGGSDHPK 218

RESULT 5
 AAR72793
 ID AAR72793 standard; Protein; 232 AA.

AC AAR72793;
 XX 16-OCT-1995 (first entry)
 DT
 DE Glutathione-s-transferase.
 XX
 KW Scl-1; senescent cell-derived inhibitor; DNA synthesis;
 KM senescence; cell proliferation; cancer; therapeutic; vulnary;
 KW fusion protein; glutathione-s-transferase.
 XX Schistosoma japonicum.
 OS
 XX
 PN WO9506415-A.
 PD
 XX 09-MAR-1995.
 PF
 XX 26-AUG-1994; 94WO-US09700.
 PR
 XX 13-JUL-1994; 94US-0274535.
 PR 30-AUG-1993; 93US-0113372.
 PR 17-NOV-1993; 93US-0153564.
 PR 03-JAN-1994; 94US-0160814.
 PR 25-FEB-1994; 94US-0203535.
 PR 15-APR-1994; 94US-0229420.
 PR 30-JUN-1994; 94US-0268439.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PI Smith JR;
 XX
 DR WPI; 1995-131101/17.
 DR N-PSDB; AAQ86774.
 XX
 PT Nucleic acid encoding a protein or polypeptide that inhibits DNA
 XX synthesis in a recipient cell - useful to inhibit cell

PT proliferation in tumour cells, treat wound or burn tissue, or as
PT an antiviral or antiparasitic agent
XX
XX
PS Disclosure; Page 130; 169pp; English.
XX
CC The senescent cell-derived inhibitor-1 (SDI-1) inhibits DNA
CC synthesis in a recipient cell and is capable of associating with a
CC cyclin (esp. cyclin D1) or cyclin-dependent kinase (esp. CDK2).
CC Production of a fusion protein of SDI-1 and glutathione-S-transferase
CC (AAR72793) facilitated the preparation and isolation of recombinant
CC SDI-1.
CC
XX
SQ Sequence 232 AA;
Query Match 100.0%; Score 1179; DB 16; Length 232;
Best Local Similarity 100.0%; Pred. No. 4.3e-113;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLRYRDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLRYRDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVLTQSMALIRYIADKHNMLGCGPKERAEISMLEGAVLDIRYGVSRIVASKDFELTKV 120
DB 61 GDVLTQSMALIRYIADKHNMLGCGPKERAEISMLEGAVLDIRYGVSRIVASKDFELTKV 120
QY 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPNCCLDAFPKLYCFK 180
DB 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPNCCLDAFPKLYCFK 180
QY 181 KRIEAIPOIDKYLKSSKXIAMPLOGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKXIAMPLOGWQATFGGDHPPK 218
RESULT 6
ID AAW73909 standard; Protein: 240 AA.
AC AAW73909;
XX
XX AAW73909;
XX
DT 13-APR-1999 (first entry)
DE GST protein sequence.
XX
XX
KM Cell surface nonexpressive functional protein; cell-permeating protein;
KM protein screening; protein production; GST; glutathione-S-transferase.
XX
XX Schistosoma japonicum.
OS
XX JP11009278-A.
PN
XX 19-JAN-1999.
PD
XX 23-JUN-1997; 97JP-0165788.
PF
XX 23-JUN-1997; 97JP-0165788.
PR
XX 23-JUN-1997; 97JP-0165788.
XX
XX (SUME) SUMITOMO ELECTRIC IND CO.
XX
XX WPI; 1999-145895/13.
DR
XX N-PSDB; AAX01312.
XX
XX Microbe having a nonexpressive functional protein on its surface -
XX PT for screening and large-scale preparation of a selected function
XX PT protein
XX
XX Disclosure; Page 8-9; 13pp; Japanese.
XX
XX This sequence is the glutathione-S-transferase (GST) protein, which
XX CC was expressed in the microbe of the invention. The microbe has a cell
XX CC surface nonexpressive functional protein on its surface, fused to a
XX CC support protein derived from a cell-permeating protein. The microbe can

CC be used for screening for biological activity of a cell surface
CC nonexpressive protein. It can also be used for preparation of a selected
CC functional protein in a large amounts.
XX
XX
SQ Sequence 240 AA;
Query Match 100.0%; Score 1179; DB 20; Length 240;
Best Local Similarity 100.0%; Pred. No. 4.6e-113;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLRYRDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLRYRDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVLTQSMALIRYIADKHNMLGCGPKERAEISMLEGAVLDIRYGVSRIVASKDFELTKV 120
DB 61 GDVLTQSMALIRYIADKHNMLGCGPKERAEISMLEGAVLDIRYGVSRIVASKDFELTKV 120
QY 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPNCCLDAFPKLYCFK 180
DB 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPNCCLDAFPKLYCFK 180
QY 181 KRIEAIPOIDKYLKSSKXIAMPLOGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKXIAMPLOGWQATFGGDHPPK 218
RESULT 7
ID AAB31612 standard; Protein: 245 AA.
AC AAB31612;
XX
XX AAB31612;
XX
DT 30-APR-2001 (first entry)
DE Amino acid sequence of glutathione-S-transferase (GST).
XX
XX
XX Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;
XX lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;
XX glutathione-S-transferase; GST.
XX
XX Unidentified.
OS
XX W0200104344-A2.
PN
XX 18-JAN-2001.
PD
XX 10-JUL-2000; 2000WO-US18828.
PF
XX 08-JUL-1999; 99US-0143757.
PR
XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
PA
XX Siegel M, Chu NR, Mizzen LA;
PI
XX WPI; 2001-138361/14.
DR
XX N-PSDB; AAF25015.
XX
XX Screening for compounds that stimulate Th1-like responses in CD4+ T
XX PT lymphocyte cells -
XX PT lymphocyte cells
XX
XX Example 9; Fig 8; 88pp; English.
XX
XX The present sequence represents a glutathione-S-transferase (GST), which
XX CC was used linked to HPV16 E7 protein, for purification purposes. The E7
XX CC protein was used to construct a fusion protein with Mycobacterium bovis
XX CC heat shock proteins (Hsp). The fusion proteins are used in the method of
XX CC the invention. The specification describes a method of determining
XX CC whether a compound stimulates a Th1-like response. Th1 cells are a subset
XX CC of CD4+ T lymphocyte cells. The method comprises contacting native
XX CC lymphocytes in vitro with a fusion protein comprising at least a fragment
XX CC of Hsp, and then detecting the Th1-like response exhibited by the cell
XX CC sample. The proteins which may be used in the method of the invention are

CC Hsp65, Hsp40, Hsp10, Hsp60, and Hsp71. The method may be used to identify
 CC compounds that stimulate Th1-like responses in response to microbial
 CC pathogens.
 CC
 XX
 SQ Sequence 245 AA;
 Query Match 100.0%; Score 1179; DB 22; Length 245;
 Best Local Similarity 100.0%; Pred. No. 4.7e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKTQSMATIRYIADKHNMLGCGCPKERAISMLBGAVIDIRYGSRIYASKDFETLKY 120
 DB 61 GDVKTQSMATIRYIADKHNMLGCGCPKERAISMLBGAVIDIRYGSRIYASKDFETLKY 120
 QY 121 DFLSKLPENLKMFEEDRLCHHTYTLNGDHYTHPDMFLYDALDVLYLMDPMDAPFKLYCFK 180
 DB 121 DFLSKLPENLKMFEEDRLCHHTYTLNGDHYTHPDMFLYDALDVLYLMDPMDAPFKLYCFK 180
 QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGSDHPK 218
 RESULT 8
 AAM74114
 ID AAM74114 standard; peptide: 247 AA.
 AC AAM74114;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
 KM D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Mahony DJ, Patterson CA, Singleton J;
 XX
 DR WPI: 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example: Page 217; 294pp; English.
 CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (i) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),

CC hPEPT1, D2H and human sucrose-isomaltase complex (hsi). (i) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (i) may also provide targeting to the GI tract. Other uses
 CC of (i) are: (1) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (i).
 CC Immunogenic analogues or derivatives of (i) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (i), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 XX
 SQ Sequence 247 AA;
 Query Match 100.0%; Score 1179; DB 20; Length 247;
 Best Local Similarity 100.0%; Pred. No. 4.7e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKTQSMATIRYIADKHNMLGCGCPKERAISMLBGAVIDIRYGSRIYASKDFETLKY 120
 DB 61 GDVKTQSMATIRYIADKHNMLGCGCPKERAISMLBGAVIDIRYGSRIYASKDFETLKY 120
 QY 121 DFLSKLPENLKMFEEDRLCHHTYTLNGDHYTHPDMFLYDALDVLYLMDPMDAPFKLYCFK 180
 DB 121 DFLSKLPENLKMFEEDRLCHHTYTLNGDHYTHPDMFLYDALDVLYLMDPMDAPFKLYCFK 180
 QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGSDHPK 218
 RESULT 9
 AAM74120
 ID AAM74120 standard; peptide: 247 AA.
 AC AAM74120;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
 KM D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Mahony DJ, Patterson CA, Singleton J;
 XX
 DR WPI: 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the

PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastro-intestinal tract, e.g. insulin or leuprolide
 XX
 XX
 PS Example: Page 221; 294pp; English.

CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX

SO Sequence 247 AA;

Query Match 100.0%; Score 1179; DB 20; Length 247;
 Best Local Similarity 100.0%; Pred. No. 4.7e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGKWKIKGLVPTLLLEYLEEKYEENLYERDEGDKMNKKFELGLEPNNLPYYID 60
 DB 1 MSPILGKWKIKGLVPTLLLEYLEEKYEENLYERDEGDKMNKKFELGLEPNNLPYYID 60
 QY 61 GPKVLTOSMAIIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYVSRIATSKDEFTLV 120
 DB 61 GPKVLTOSMAIIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYVSRIATSKDEFTLV 120
 QY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYNDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYNDPCLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218

RESULT 10

ID AAW74103 standard; peptide: 247 AA.

AC AAW74103;

DT 04-MAY-1999 (first entry)

DE GST/GI transport receptor binding protein fusion sequence.

KM Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KM D2H; hPPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

PN W09851325-A2.

PD 19-NOV-1998.

PF 15-MAY-1998; 98WO-US10088.

XX

PR 15-MAY-1997; 97US-0046595.
 XX
 XX (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IO;
 PI Omanony DJ, Patterson CA, Singleton J;
 DR WPI, 1999-009568/01.

PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastro-intestinal tract, e.g. insulin or leuprolide
 XX
 XX
 PS Example: Page 209; 294pp; English.

CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX

SO Sequence 247 AA;

Query Match 100.0%; Score 1179; DB 20; Length 247;
 Best Local Similarity 100.0%; Pred. No. 4.7e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGKWKIKGLVPTLLLEYLEEKYEENLYERDEGDKMNKKFELGLEPNNLPYYID 60
 DB 1 MSPILGKWKIKGLVPTLLLEYLEEKYEENLYERDEGDKMNKKFELGLEPNNLPYYID 60
 QY 61 GPKVLTOSMAIIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYVSRIATSKDEFTLV 120
 DB 61 GPKVLTOSMAIIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYVSRIATSKDEFTLV 120
 QY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYNDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYNDPCLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218

RESULT 11

ID AAW74112 standard; peptide: 248 AA.

AC AAW74112;

DT 04-MAY-1999 (first entry)

DE GST/GI transport receptor binding protein fusion sequence.

KM Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KM D2H; hPPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;

KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX Homo sapiens.
 OS MO9851325-A2.
 PM 19-NOV-1998.
 PD 15-MAY-1998; 98WO-US10088.
 XX 15-MAY-1997; 97US-0046595.
 PR 15-MAY-1997; 97US-0046595.
 XX (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Omahony DJ, Patterson CA, Singleton J;
 XX WPI: 1999-009568/01.
 DR
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example: Page 215-216; 294pp; English.
 XX
 CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPRT1),
 CC hPRT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX
 SO Sequence 248 AA:
 Query Match 100.0%; Score 1179; DB 20; Length 248;
 Best Local Similarity 100.0%; Pred. No. 4.8e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MSPILGYWKIKGLVOPTRLLEYLEKYEHEHLERDEGDKWRKKFELGIEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPTRLLEYLEKYEHEHLERDEGDKWRKKFELGIEFPNLPYYID 60
 OY 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRIRAYSDPFTLVK 120
 DB 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRIRAYSDPFTLVK 120
 OY 121 DFLSKLPKMLKMFEDRLCKRTYVNGDVTHPDMLYDALDVLYMDPMLDAEPKLVCFK 180
 DB 121 DFLSKLPKMLKMFEDRLCKRTYVNGDVTHPDMLYDALDVLYMDPMLDAEPKLVCFK 180
 OY 181 KRTEALPQIDKYLKSSKRYTAMPLOGMATFGGSDHPK 218
 DB 181 KRTEALPQIDKYLKSSKRYTAMPLOGMATFGGSDHPK 218

ID AAM74113 standard; peptide: 248 AA.
 XX
 AC AAM74113;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein; hSI; hPRT1;
 KM D2H; hPRT1; human; GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX
 XX Homo sapiens.
 OS MO9851325-A2.
 PM 19-NOV-1998.
 PD 15-MAY-1998; 98WO-US10088.
 XX 15-MAY-1997; 97US-0046595.
 PR 15-MAY-1997; 97US-0046595.
 XX (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Omahony DJ, Patterson CA, Singleton J;
 XX WPI: 1999-009568/01.
 DR
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example: Page 216; 294pp; English.
 XX
 CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPRT1),
 CC hPRT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX
 SO Sequence 248 AA:
 Query Match 100.0%; Score 1179; DB 20; Length 248;
 Best Local Similarity 100.0%; Pred. No. 4.8e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MSPILGYWKIKGLVOPTRLLEYLEKYEHEHLERDEGDKWRKKFELGIEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPTRLLEYLEKYEHEHLERDEGDKWRKKFELGIEFPNLPYYID 60
 OY 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRIRAYSDPFTLVK 120
 DB 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRIRAYSDPFTLVK 120

QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPMLDAFPKLVCFK 180
 DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPMLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKXIAMPLOGMQATFGGDDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKXIAMPLOGMQATFGGDDHPK 218

RESULT 13

AAW74121
 ID AAW74121 standard; peptide: 248 AA.

AAW74121;

04-MAY-1999 (first entry)

GST/GI transport receptor binding protein fusion sequence.

XX Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

XX Homo sapiens.

XX MO9851325-A2.

XX 19-NOV-1998.

XX 15-MAY-1998; 98WO-US10088.

XX 15-MAY-1997; 97US-0046595.

XX (CYTO-) CYTOGEN CORP.

XX (ELAN-) ELAN CORP PLC.

XX Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ,
 PI Omahony DJ, Patterson CA, Singleton J;

XX WPI; 1999-009568/01.

XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide

XX Example; Page 222; 294pp; English.

XX This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

XX Sequence 248 AA;

XX Query Match 100.0%; Score 1179; DB 20; Length 248;

Best Local Similarity 100.0%; Pred. No. 4 8e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYNKIGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEPPLPYID 60

DB 1 MSPILGYNKIGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEPPLPYID 60

QY 61 GDVKLTQSMATIRYADHNMLGGCPKRAEISMLEGAVLDIRGVSRINASKDEETLKV 120

DB 61 GDVKLTQSMATIRYADHNMLGGCPKRAEISMLEGAVLDIRGVSRINASKDEETLKV 120

QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPMLDAFPKLVCFK 180

DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPMLDAFPKLVCFK 180

QY 181 KRIEAIPOIDKYLKSSKXIAMPLOGMQATFGGDDHPK 218

DB 181 KRIEAIPOIDKYLKSSKXIAMPLOGMQATFGGDDHPK 218

RESULT 14

AAW74122
 ID AAW74122 standard; peptide: 248 AA.

AAW74122;

04-MAY-1999 (first entry)

GST/GI transport receptor binding protein fusion sequence.

XX Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

XX Homo sapiens.

XX MO9851325-A2.

XX 19-NOV-1998.

XX 15-MAY-1998; 98WO-US10088.

XX 15-MAY-1997; 97US-0046595.

XX (CYTO-) CYTOGEN CORP.

XX (ELAN-) ELAN CORP PLC.

XX Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ,
 PI Omahony DJ, Patterson CA, Singleton J;

XX WPI; 1999-009568/01.

XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide

XX Example; Page 222-223; 294pp; English.

XX This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents

CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

XX Sequence 248 AA:

Query Match 100.0%; Score 1179; DB 20; Length 248;

Best Local Similarity 100.0%; Pred. No. 4.8e-113; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPRLLEYLEEYEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPRLLEYLEEYEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTOSMAIIRYADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKY 120
 DB 61 GDVKLTOSMAIIRYADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKY 120
 QY 121 DFLSKLPKMKFEDRLCHRTYLNQDHTHPDFMLYDALDVLYMDPCLDAEPKLVCFK 180
 DB 121 DFLSKLPKMKFEDRLCHRTYLNQDHTHPDFMLYDALDVLYMDPCLDAEPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKRYIAMPLOGWQATFGGDDHPRK 218
 DB 181 KRIEAIPOIDKYLKSSKRYIAMPLOGWQATFGGDDHPRK 218

RESULT 15

AAW74100 AAW74100 standard; peptide; 248 AA.

XX AAW74100;

DT 04-MAY-1999 (first entry)

DE GST/GI transport receptor binding protein fusion sequence.

KM Gastro-intestinal transport receptor; binding protein; hSI; HPT1;

KM DH: hPEPT1; human: GI tract receptor; sucrose-isomaltase complex;

KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;

KM therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

PN W09851325-A2.

PD 19-NOV-1998.

PF 15-MAY-1998; 98WO-US10088.

PR 15-MAY-1997; 97US-0046595.

PA (CYTO-) CYTOGEN CORP.

PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JF;

DR WPI; 1999-009568/01.

XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX Example: Page 207; 294pp; English.
 CC This sequence represents a fusion protein between

CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

XX Sequence 248 AA:

Query Match 100.0%; Score 1179; DB 20; Length 248;

Best Local Similarity 100.0%; Pred. No. 4.8e-113; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPRLLEYLEEYEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPRLLEYLEEYEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTOSMAIIRYADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKY 120
 DB 61 GDVKLTOSMAIIRYADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKY 120
 QY 121 DFLSKLPKMKFEDRLCHRTYLNQDHTHPDFMLYDALDVLYMDPCLDAEPKLVCFK 180
 DB 121 DFLSKLPKMKFEDRLCHRTYLNQDHTHPDFMLYDALDVLYMDPCLDAEPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKRYIAMPLOGWQATFGGDDHPRK 218
 DB 181 KRIEAIPOIDKYLKSSKRYIAMPLOGWQATFGGDDHPRK 218

Search completed: May 20, 2003, 11:33:54
 Job time: 18.4058 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 7.02587 Seconds

(without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179
Sequence: 1 MSPILGYWKIKGLVQPTRL.....IAMPLOGMQATFGCGDHPK 218Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1179	100.0	218	1	US-08-470-837-26
2	1179	100.0	218	4	US-08-868-452-26
3	1179	100.0	231	4	US-08-129-722A-2
4	1179	100.0	232	4	US-08-337-874A-6
5	1179	100.0	232	5	PCT-US94-09700-6
6	1179	100.0	272	4	US-08-910-820-6
7	1179	100.0	272	4	US-08-910-820-6
8	1179	100.0	282	4	US-08-910-820-3
9	1179	100.0	282	4	US-08-910-820-5
10	1179	100.0	307	2	US-08-216-894-6
11	1179	100.0	307	4	US-09-115-746-6
12	1179	100.0	331	4	US-09-217-228-6
13	1179	100.0	352	1	US-08-395-507-1
14	1179	100.0	354	2	US-08-216-894-4
15	1179	100.0	354	4	US-09-115-746-4
16	1179	100.0	362	1	US-08-385-507-2
17	1179	100.0	367	4	US-08-337-874A-11
18	1179	100.0	397	5	PCT-US94-09700-11
19	1179	100.0	412	4	US-09-366-009-34
20	1179	100.0	422	4	US-09-217-228-7
21	1179	100.0	436	3	US-08-737-248-4
22	1179	100.0	435	5	PCT-US93-04439-1
23	1179	100.0	439	4	US-08-506-2968-67
24	1179	100.0	442	4	US-08-506-2968-70
25	1179	100.0	443	4	US-08-506-2968-76
26	1179	100.0	447	4	US-08-506-2968-73
27	1179	100.0	472	2	US-08-216-894-10

28	1179	100.0	472	4	US-09-115-746-10	Sequence 10, Appl
29	1179	100.0	514	4	US-08-974-549A-605	Sequence 605, App
30	1179	100.0	515	4	US-08-974-549A-604	Sequence 604, App
31	1179	100.0	517	4	US-08-974-549A-606	Sequence 606, App
32	1179	100.0	530	4	US-08-974-549A-603	Sequence 603, App
33	1179	100.0	536	3	US-08-974-180-20	Sequence 20, Appl
34	1179	100.0	538	4	US-08-974-549A-602	Sequence 602, App
35	1179	100.0	542	4	US-08-506-2968-69	Sequence 69, Appl
36	1179	100.0	545	4	US-08-506-2968-75	Sequence 75, Appl
37	1179	100.0	547	4	US-08-506-2968-72	Sequence 72, Appl
38	1179	100.0	559	4	US-08-506-2968-66	Sequence 66, Appl
39	1179	100.0	564	2	US-08-216-894-2	Sequence 2, Appl
40	1179	100.0	564	4	US-09-115-746-2	Sequence 2, Appl
41	1179	100.0	579	2	US-08-864-224-11	Sequence 11, Appl
42	1179	100.0	632	4	US-08-506-2968-74	Sequence 74, Appl
43	1179	100.0	635	4	US-08-506-2968-71	Sequence 71, Appl
44	1179	100.0	643	2	US-08-216-894-8	Sequence 8, Appl
45	1179	100.0	643	4	US-09-115-746-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-470-837-26
; Sequence 26, Application US/08470837
; Patent No. 5800811

GENERAL INFORMATION:

APPLICANT: Nimm, Marcel E.
APPLICANT: Tuan, Frederick L.
APPLICANT: Wu, Lingtao
APPLICANT: Cheung, David T.
TITLE OF INVENTION: Transforming Growth Factor B Fusion
TITLE OF INVENTION: and
TITLE OF INVENTION: Their Use in Wound Healing
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 1150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025-3395

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,837
FILING DATE:
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Sharp, Janice A.
REGISTRATION NUMBER: 34,051
REFERENCE/DOCKET NUMBER: 30630-1US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-837-26

Query Match 100.0% Score 1179; DB 1; Length 218;
Best Local Similarity 100.0% Pred. No. 2.3e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVQPTRLLELYEYERHLYERDEGDKMKNKFFELGLEPNTLPYYID 60

```
Db 1 MSPIIGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
      |||
Qy 61 GDVKLTQSMALIRYIADKHNMLGCGCPKERAISMLEGAVLDIRYGVSRIVSKDFETLKV 120
      |||
Db 61 GDVKLTQSMALIRYIADKHNMLGCGCPKERAISMLEGAVLDIRYGVSRIVSKDFETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPWCDAFPKLVCFK 180
      |||
Db 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPWCDAFPKLVCFK 180
Qy 181 KRIFAIPQIDKYLKSSKXYIAMPLOGWQATFGGDDHPK 218
      |||
Db 181 KRIFAIPQIDKYLKSSKXYIAMPLOGWQATFGGDDHPK 218

RESULT 2
US-08-452-26
: Sequence 26, Application US/08868452C
: Patent No. 6352972
: GENERAL INFORMATION:
: APPLICANT: Marcel E. Nimmi
: APPLICANT: Frederick L. Hall
: APPLICANT: Lingtao Wu
: APPLICANT: Edwin Shors
: TITLE OF INVENTION: BONE MORPHOGENETIC PROTEINS AND THEIR
: FILE REFERENCE: 1/972-11
: CURRENT APPLICATION NUMBER: US/08/868,452C
: NUMBER OF SEQ ID NOS: 51
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 26
: LENGTH: 218
: TYPE: PRT
: ORGANISM: Human
US-08-868-452-26

Query Match      100.0%; Score 1179; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.3e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSPIIGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
      |||
Db 1 MSPIIGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
Qy 61 GDVKLTQSMALIRYIADKHNMLGCGCPKERAISMLEGAVLDIRYGVSRIVSKDFETLKV 120
      |||
Db 61 GDVKLTQSMALIRYIADKHNMLGCGCPKERAISMLEGAVLDIRYGVSRIVSKDFETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPWCDAFPKLVCFK 180
      |||
Db 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPWCDAFPKLVCFK 180
Qy 181 KRIFAIPQIDKYLKSSKXYIAMPLOGWQATFGGDDHPK 218
      |||
Db 181 KRIFAIPQIDKYLKSSKXYIAMPLOGWQATFGGDDHPK 218

RESULT 3
US-08-129-722A-2
: Sequence 2, Application US/08129722A
: Patent No. 6303369
: GENERAL INFORMATION:
: APPLICANT: Spana, Carl
: APPLICANT: Fargnoli, Joseph
: APPLICANT: Bolen, Joseph B.
: TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burton Rodney
: STREET: P. O. Box 4000
```

```
      CITY: Princeton
      STATE: New Jersey
      COUNTRY: U.S.A.
      ZIP: 08543-4000
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patent Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/129,722A
      FILING DATE:
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
      NAME: Gaul, Timothy J.
      REGISTRATION NUMBER: 33,111
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (609) 252-5901
      TELEFAX: (609) 252-4526
      INFORMATION FOR SEQ ID NO: 2:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 231 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      MOLECULE TYPE: protein
US-08-129-722A-2

Query Match      100.0%; Score 1179; DB 4; Length 231;
Best Local Similarity 100.0%; Pred. No. 2.5e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSPIIGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
      |||
Db 1 MSPIIGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
Qy 61 GDVKLTQSMALIRYIADKHNMLGCGCPKERAISMLEGAVLDIRYGVSRIVSKDFETLKV 120
      |||
Db 61 GDVKLTQSMALIRYIADKHNMLGCGCPKERAISMLEGAVLDIRYGVSRIVSKDFETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPWCDAFPKLVCFK 180
      |||
Db 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPWCDAFPKLVCFK 180
Qy 181 KRIFAIPQIDKYLKSSKXYIAMPLOGWQATFGGDDHPK 218
      |||
Db 181 KRIFAIPQIDKYLKSSKXYIAMPLOGWQATFGGDDHPK 218

RESULT 4
US-08-327-874A-6
: Sequence 6, Application US/08327874A
: Patent No. 6372249
: GENERAL INFORMATION:
: APPLICANT: BAYLOR COLLEGE OF MEDICINE
: APPLICANT: SMITH, JAMES R.
: APPLICANT: DRUTZ, DAVID J.
: APPLICANT: WILSON, DEBORAH R.
: APPLICANT: ZUMSTEIN, LOUIS A.
: TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ROGERS & WELLS
: STREET: 200 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10166
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,874A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT US94/09700
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/327,874
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: NORTON, GERARD P.
REGISTRATION NUMBER: 36,621
REFERENCE/DOCKET NUMBER: 3634-8-CIP10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 878-3148
TELEFAX: (212) 878-8375
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Schistosoma japonicum
IMMEDIATE SOURCE:
CLONE: GST
US-08-327-874A-6

Query Match 100.0%; Score 1179; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 2,66-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVPTLLLEYLEEKYEHLHYERDEGDKWNNKFFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLLEYLEEKYEHLHYERDEGDKWNNKFFELGLEFPNLPYYID 60
QY 61 GDVKTLSMAITRYADKHNMLGGCPKERAETISMLEGAVLDIRYGSRIAYSKDEFTLKV 120
DB 61 GDVKTLSMAITRYADKHNMLGGCPKERAETISMLEGAVLDIRYGSRIAYSKDEFTLKV 120
QY 121 DFLSLTPMLKWFEDRLCHKTLYNDGHTHPDNLXDADLVLYLNDPKLDAFPLVYCK 180
DB 121 DFLSLTPMLKWFEDRLCHKTLYNDGHTHPDNLXDADLVLYLNDPKLDAFPLVYCK 180
QY 181 KRIEATPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218
DB 181 KRIEATPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218

RESULT 5
PCT-US94-09700-6
Sequence 6, Application PC/TUS9409700
GENERAL INFORMATION:
APPLICANT: BAYLOR COLLEGE OF MEDICINE
APPLICANT: SMITH, JAMES R.
TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
TITLE OF INVENTION: DNA SYNTHESIS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSER: HORREY & SIMON
STREET: 1299 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09700
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: AUERBACH, JEFFREY I.
REGISTRATION NUMBER: 32,680
REFERENCE/DOCKET NUMBER: 225-102-CIP7-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 383-7451
TELEFAX: (202) 383-6610
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Schistosoma japonicum
IMMEDIATE SOURCE:
CLONE: GST
PCT-US94-09700-6

Query Match 100.0%; Score 1179; DB 5; Length 232;
Best Local Similarity 100.0%; Pred. No. 2,66-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVPTLLLEYLEEKYEHLHYERDEGDKWNNKFFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLLEYLEEKYEHLHYERDEGDKWNNKFFELGLEFPNLPYYID 60

QY 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
DB 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
QY 121 DFLSKPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVVLYNDPMLDAFPKLVCFK 180
DB 121 DFLSKPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVVLYNDPMLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPPK 218

RESULT 6
US-08-910-820-4

; Sequence 4, Application US/08910820
; Patent No. 6258579
; GENERAL INFORMATION:
; APPLICANT: Mercurio, Frank
; APPLICANT: Zhu, Hengyi
; APPLICANT: Barbosa, Miguel
; APPLICANT: Li, Gfan
; APPLICANT: Murray, Brian W.
; TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
; TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,820
; FILING DATE: 12-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.413C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-910-820-4

Query Match 100.0%; Score 1179; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 3.2e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKMRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKMRNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
DB 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
QY 121 DFLSKPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVVLYNDPMLDAFPKLVCFK 180
DB 121 DFLSKPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVVLYNDPMLDAFPKLVCFK 180

QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPPK 218

RESULT 7
US-08-910-820-6

; Sequence 6, Application US/08910820
; Patent No. 6258579
; GENERAL INFORMATION:
; APPLICANT: Mercurio, Frank
; APPLICANT: Zhu, Hengyi
; APPLICANT: Barbosa, Miguel
; APPLICANT: Li, Gfan
; APPLICANT: Murray, Brian W.
; TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
; TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,820
; FILING DATE: 12-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.413C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-910-820-6

Query Match 100.0%; Score 1179; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 3.2e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKMRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKMRNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
DB 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
QY 121 DFLSKPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVVLYNDPMLDAFPKLVCFK 180
DB 121 DFLSKPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVVLYNDPMLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPPK 218

RESULT 8
US-08-910-820-3

Sequence 3, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-3

Query Match 100.0%; Score 1179; DB 4; Length 282;
Best Local Similarity 100.0%; Pred. No. 3.3e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEKYEHEHLYERDEGDKWRKKFELGLEFNNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEKYEHEHLYERDEGDKWRKKFELGLEFNNLPYYID 60
QY 61 GDVKLTQSMAIIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGVSRIVASDFFELKY 120
DB 61 GDVKLTQSMAIIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGVSRIVASDFFELKY 120
QY 121 DFLSKLPENLKMFEEDRLCHRTYLNQDHYTHPDMLTDALDVLVYMPMCLDAPPKLVCFK 180
DB 121 DFLSKLPENLKMFEEDRLCHRTYLNQDHYTHPDMLTDALDVLVYMPMCLDAPPKLVCFK 180
QY 181 KRTEAIPOIDKYLKSSKIYAMPLOGQWATFGGSDHPPK 218
DB 181 KRTEAIPOIDKYLKSSKIYAMPLOGQWATFGGSDHPPK 218

RESULT 9
US-08-910-820-5
Sequence 5, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-5

Query Match 100.0%; Score 1179; DB 4; Length 282;
Best Local Similarity 100.0%; Pred. No. 3.3e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEKYEHEHLYERDEGDKWRKKFELGLEFNNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEKYEHEHLYERDEGDKWRKKFELGLEFNNLPYYID 60
QY 61 GDVKLTQSMAIIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGVSRIVASDFFELKY 120
DB 61 GDVKLTQSMAIIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGVSRIVASDFFELKY 120
QY 121 DFLSKLPENLKMFEEDRLCHRTYLNQDHYTHPDMLTDALDVLVYMPMCLDAPPKLVCFK 180
DB 121 DFLSKLPENLKMFEEDRLCHRTYLNQDHYTHPDMLTDALDVLVYMPMCLDAPPKLVCFK 180
QY 181 KRTEAIPOIDKYLKSSKIYAMPLOGQWATFGGSDHPPK 218
DB 181 KRTEAIPOIDKYLKSSKIYAMPLOGQWATFGGSDHPPK 218

RESULT 10
US-08-216-894-6
Sequence 6, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-6

Query Match          100.0%; Score 1179; DB 2; Length 307;
Best Local Similarity 100.0%; Pred. No. 3.8e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRGVSRIVASKDEFTLV 120
DB 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRGVSRIVASKDEFTLV 120
QY 121 DFLSLPEMLKMFEDRLCHKTYLNGDHYTHPDPMLYDALDVLVLYMDPMCLDAFPKLVCFK 180
DB 121 DFLSLPEMLKMFEDRLCHKTYLNGDHYTHPDPMLYDALDVLVLYMDPMCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDDHPK 218

RESULT 11
US-09-115-746-6
Sequence 6, Application US/09115746
Patent No. 6228601
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAPOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,746
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,894
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
```

```
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-6

Query Match          100.0%; Score 1179; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 3.8e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRGVSRIVASKDEFTLV 120
DB 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRGVSRIVASKDEFTLV 120
QY 121 DFLSLPEMLKMFEDRLCHKTYLNGDHYTHPDPMLYDALDVLVLYMDPMCLDAFPKLVCFK 180
DB 121 DFLSLPEMLKMFEDRLCHKTYLNGDHYTHPDPMLYDALDVLVLYMDPMCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDDHPK 218

RESULT 12
US-09-217-228-6
Sequence 6, Application US/09217228
Patent No. 6323178
GENERAL INFORMATION:
APPLICANT: Butler, Jon P.
APPLICANT: Hale, John E.
APPLICANT: Heath Jr., William F.
APPLICANT: Schoner, Brigitte E.
APPLICANT: Helman, Mark L.
APPLICANT: Becker, Gerald W.
APPLICANT: Varshavsky, Alexander D.
TITLE OF INVENTION: Beta-lipotropin and Uses Thereof
FILE REFERENCE: X-12139
CURRENT APPLICATION NUMBER: US/09/217,228
CURRENT FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 35
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 6
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GST fusion
OTHER INFORMATION: partner
US-09-217-228-6

Query Match          100.0%; Score 1179; DB 4; Length 331;
Best Local Similarity 100.0%; Pred. No. 4.2e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRGVSRIVASKDEFTLV 120
DB 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRGVSRIVASKDEFTLV 120
```


QY 121 DFLSKLPEMLKMFEDRLCHRTYLNQDVTHTPDMLYDALDVLVYMDPCLDAFPAKLVCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHRTYLNQDVTHTPDMLYDALDVLVYMDPCLDAFPAKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWATFGGSDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWATFGGSDHPPK 218

RESULT 13

US-08-395-507-1
Sequence 1, Application US/08395507
Patent No. 5578456
GENERAL INFORMATION:
APPLICANT: Fujimura, Katsuya
APPLICANT: Ueno, Eiichi
APPLICANT: Fujii, No. 5578456uyuk1
APPLICANT: Okada, Masahisa
TITLE OF INVENTION: Anti-treponema Pallidum Antibody
TITLE OF INVENTION: Immunoassay
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pollock, Vande Sande & Priddy
STREET: 1990 M Street, NW, Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,507
FILING DATE: 27-FEB-1995
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP Hei-6-54672
FILING DATE: 28-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Amerjick, Burton A.
REGISTRATION NUMBER: 24,852
REFERENCE/DOCKET NUMBER: 1581/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 331-7111
TELEFAX: (202) 293-6229
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: rabbit
FEATURE:
NAME/KEY: Protein
LOCATION: 1..352
OTHER INFORMATION: /note="G15 Antigen"

Query Match

Best Local Similarity 100.0%; Score 1179; DB 1; Length 352;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGIEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGIEFPNLPYYID 60
QY 61 GDVKTOSMAIIRIYIADKHNMLGGCPKERAISMLEGAVLDIRGVSRIVASKDFETLVK 120

DB 61 GDVKTOSMAIIRIYIADKHNMLGGCPKERAISMLEGAVLDIRGVSRIVASKDFETLVK 120
QY 121 DFLSKLPEMLKMFEDRLCHRTYLNQDVTHTPDMLYDALDVLVYMDPCLDAFPAKLVCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHRTYLNQDVTHTPDMLYDALDVLVYMDPCLDAFPAKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWATFGGSDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWATFGGSDHPPK 218

RESULT 14

US-08-216-894-4
Sequence 4, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-4

Query Match

Best Local Similarity 100.0%; Score 1179; DB 2; Length 354;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGIEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGIEFPNLPYYID 60
QY 61 GDVKTOSMAIIRIYIADKHNMLGGCPKERAISMLEGAVLDIRGVSRIVASKDFETLVK 120
DB 61 GDVKTOSMAIIRIYIADKHNMLGGCPKERAISMLEGAVLDIRGVSRIVASKDFETLVK 120
QY 121 DFLSKLPEMLKMFEDRLCHRTYLNQDVTHTPDMLYDALDVLVYMDPCLDAFPAKLVCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHRTYLNQDVTHTPDMLYDALDVLVYMDPCLDAFPAKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWATFGGSDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWATFGGSDHPPK 218

RESULT 15
US-09-115-746-4
Sequence 4, Application US/09115746
Patent No. 6228601
GENERAL INFORMATION:
APPLICANT: Kirchoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,746
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,894
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5359
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-4

Query Match 100.0%; Score 1179; DB 4; Length 354;
Best Local Similarity 100.0%; Pred. No. 4: 6e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVOPTRILLEYLEEKYEHLIERDEGDKWNRKKFELGLEFPNLPYIID 60
DB 1 MSPILGYWKIKGLVOPTRILLEYLEEKYEHLIERDEGDKWNRKKFELGLEFPNLPYIID 60
QY 61 GDVKLTQSMALIRIYADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIRIAYSKDFETLKV 120
DB 61 GDVKLTQSMALIRIYADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIRIAYSKDFETLKV 120
QY 121 DFLSLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYNDPMLCLDAFPKLVCFK 180
DB 121 DFLSLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYNDPMLCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSKSYIAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSKSYIAMPLOGWQATFGGDHPK 218

Search completed: May 20, 2003, 11:40:56
Job time : 8.02587 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 34.7335 Seconds

(Without alignments)
622.443 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179
Sequence: 1 MSPILGWIKIKGLVQPTRL.....IAMPLOGWATFGGDDHPK 218

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/FCI_NEM_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCRTS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1179	100.0	218	9	US-10-081-408-4
2	1179	100.0	239	10	US-09-823-153-7
3	1179	100.0	245	9	US-10-267-311-23
4	1179	100.0	272	9	US-09-844-988-4
5	1179	100.0	272	9	US-09-844-988-6
6	1179	100.0	272	10	US-09-844-908-4
7	1179	100.0	272	10	US-09-844-908-6
8	1179	100.0	282	9	US-09-844-988-3
9	1179	100.0	282	9	US-09-844-988-5
10	1179	100.0	282	10	US-09-844-908-3
11	1179	100.0	282	10	US-09-844-908-5
12	1179	100.0	298	9	US-09-910-600-25
13	1179	100.0	324	9	US-10-267-311-25
14	1179	100.0	331	9	US-09-824-438-6
15	1179	100.0	348	9	US-09-910-600-22
16	1179	100.0	348	9	US-09-910-600-23
17	1179	100.0	348	9	US-09-910-600-24
18	1179	100.0	348	9	US-09-910-600-25
19	1179	100.0	354	10	US-09-823-153-8

20	1179	100.0	364	9	US-09-788-268-14	Sequence 14, Appl
21	1179	100.0	394	10	US-09-990-578-4	Sequence 4, Appl
22	1179	100.0	401	9	US-10-190-866A-1	Sequence 1, Appl
23	1179	100.0	412	9	US-09-775-964-34	Sequence 34, Appl
24	1179	100.0	422	9	US-09-824-438-7	Sequence 7, Appl
25	1179	100.0	440	9	US-09-950-634-4	Sequence 4, Appl
26	1176	99.7	649	9	US-10-174-784-9	Sequence 9, Appl
27	1149	97.5	218	9	US-10-081-408-5	Sequence 5, Appl
28	1144	97.0	988	9	US-10-081-408-20	Sequence 20, Appl
29	472.5	40.1	229	9	US-10-102-806-554	Sequence 554, App
30	187.5	15.9	222	9	US-09-784-739-4	Sequence 4, Appl
31	187.5	15.9	230	10	US-09-925-301-1318	Sequence 1318, Ap
32	171.5	14.5	222	9	US-09-784-739-3	Sequence 3, Appl
33	165.5	14.0	222	9	US-09-784-739-5	Sequence 5, Appl
34	164.5	14.0	203	9	US-09-847-208-48	Sequence 48, Appl
35	152.5	12.9	222	9	US-09-784-739-1	Sequence 1, Appl
36	99	8.4	240	9	US-09-964-899-23	Sequence 23, Appl
37	99	8.4	379	9	US-10-253-007-46	Sequence 46, Appl
38	97	8.2	233	10	US-09-765-213A-2	Sequence 2, Appl
39	94.5	8.0	412	9	US-09-214-592-31	Sequence 31, Appl
40	92	7.8	233	10	US-09-765-213A-4	Sequence 4, Appl
41	89	7.5	233	9	US-09-765-213A-6	Sequence 6, Appl
42	87.5	7.4	359	9	US-09-738-626-4895	Sequence 4895, Ap
43	83	7.0	1668	10	US-09-815-242-5654	Sequence 5654, Ap
44	83	7.0	2397	10	US-09-815-242-12265	Sequence 12265, A
45	81.5	6.9	241	9	US-09-854-133-204	Sequence 204, App

ALIGNMENTS

RESULT 1
US-10-081-408-4
Sequence 4, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Nilsen, Joakim
FILE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: US 60/272,247
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 218
TYPE: PRT
ORGANISM: Schistosoma japonicum
US-10-081-408-4

Query Match 100.0%; Score 1179; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 2e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSPILGWIKIKGLVQPTRLLEYLEKYEHLERDEGKRNKKFELGEPNNLPYTD	60
DB	1	MSPILGWIKIKGLVQPTRLLEYLEKYEHLERDEGKRNKKFELGEPNNLPYTD	60
QY	61	GVVLTQSAAIRYIADKNNMLGGCPKEAETSMLEGAVALDIRYGSRTAYSKDPETLV	120
DB	61	GVVLTQSAAIRYIADKNNMLGGCPKEAETSMLEGAVALDIRYGSRTAYSKDPETLV	120
QY	121	DELTKLPEMLKMFEDRLCKRTYLNQDHYTHPDMYDALDVLVLYMDPKLDAFPKLYCR	180
DB	121	DELTKLPEMLKMFEDRLCKRTYLNQDHYTHPDMYDALDVLVLYMDPKLDAFPKLYCR	180
QY	181	KRIEAIPOIDKYLKSSKTIYAMPLOGWATFGGDDHPK 218	
DB	181	KRIEAIPOIDKYLKSSKTIYAMPLOGWATFGGDDHPK 218	

RESULT 2
US-09-823-153-7
; Sequence 7, Application US/09823153
; Patent No. US20020025540A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; APPLICANT: Roberts, Susan
; APPLICANT: Pak, Roger
; APPLICANT: Lewis, Martin
; APPLICANT: Smith, David
; APPLICANT: Hendrick, Joseph
; APPLICANT: Vinitasky, Alexander
; TITLE OF INVENTION: ISOLATION OF FUNCTIONALLY ACTIVE GAMMA-SECRETASE PROTEIN COMPLEX
; TITLE OF INVENTION: METHODS FOR DETECTION OF ACTIVITY AND INHIBITORS THEREOF
; FILE REFERENCE: DD004
; CURRENT APPLICATION NUMBER: US/09/823.153
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 7
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Description of Artificial Sequence: PSI PEPTIDE
US-09-823-153-7
Query Match 100.0%; Score 1179; DB 10; Length 239;
Best Local Similarity 100.0%; Pred. No. 2,2e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGVKIKGLVOPRLLEYLEEKEEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGVKIKGLVOPRLLEYLEEKEEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMALIRYIADKHNLMGCGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKY 120
DB 61 GDVKLTQSMALIRYIADKHNLMGCGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKY 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPDMLYDALDVLYMDPCLDAFPPLVCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPDMLYDALDVLYMDPCLDAFPPLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOTFGGCDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOTFGGCDHPK 218
RESULT 3
US-10-267-311-23
; Sequence 23, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/10/267.311
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US/09/613.303
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143.757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 23
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATURE:
; OTHER INFORMATION: fusion sequence
US-10-267-311-23
Query Match 100.0%; Score 1179; DB 9; Length 245;
Best Local Similarity 100.0%; Pred. No. 2,3e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGVKIKGLVOPRLLEYLEEKEEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGVKIKGLVOPRLLEYLEEKEEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMALIRYIADKHNLMGCGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKY 120
DB 61 GDVKLTQSMALIRYIADKHNLMGCGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKY 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPDMLYDALDVLYMDPCLDAFPPLVCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPDMLYDALDVLYMDPCLDAFPPLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOTFGGCDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOTFGGCDHPK 218
RESULT 4
US-09-844-988-4
; Sequence 4, Application US/09844988
; Patent No. US20020158764A1
; GENERAL INFORMATION:
; APPLICANT: Mercurio, Frank
; APPLICANT: Zhu, Hengyi
; APPLICANT: Barbosa, Miguel
; APPLICANT: Li, Gian
; APPLICANT: Murray, Brian W.
; TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/844.988
; FILING DATE: 26-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/910,820
; FILING DATE: 1997-08-13
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-988-4

Query Match 100.0%; Score 1179; DB 9; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60

QY 61 GVKLTQSMAIIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRAYSKDFTLKY 120
DB 61 GVKLTQSMAIIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRAYSKDFTLKY 120

QY 121 DFLSKPEMLKMEDEDLCKHTYINGDHVTHPDMLYDALDVLYXMDPCIDAPFKLYCCK 180
DB 121 DFLSKPEMLKMEDEDLCKHTYINGDHVTHPDMLYDALDVLYXMDPCIDAPFKLYCCK 180

QY 181 KRIEAIPOIDKYLKSSKSIYAMPLOGQWATFGGSDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKSIYAMPLOGQWATFGGSDHPPK 218

RESULT 5
US-09-844-988-6
Sequence 6, Application US/09844988
Patent No. US20020158764A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gian
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,988
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/910,820
FILING DATE: 1997-08-13
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-844-988-6

Query Match 100.0%; Score 1179; DB 9; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60

QY 61 GVKLTQSMAIIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRAYSKDFTLKY 120
DB 61 GVKLTQSMAIIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRAYSKDFTLKY 120

QY 121 DFLSKPEMLKMEDEDLCKHTYINGDHVTHPDMLYDALDVLYXMDPCIDAPFKLYCCK 180
DB 121 DFLSKPEMLKMEDEDLCKHTYINGDHVTHPDMLYDALDVLYXMDPCIDAPFKLYCCK 180

QY 181 KRIEAIPOIDKYLKSSKSIYAMPLOGQWATFGGSDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKSIYAMPLOGQWATFGGSDHPPK 218

RESULT 6
US-09-844-908-4
Sequence 4, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gian
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-908-4

Query Match 100.0%; Score 1179; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60

QY 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRITAYSKDEFTLV 120
DB 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRITAYSKDEFTLV 120
QY 121 DFLSKLPEMLKAFEDRLCKHTYLNGBDHTVHPDFMLYDALDVLYMDPKLDAFPKLYCCK 180
DB 121 DFLSKLPEMLKAFEDRLCKHTYLNGBDHTVHPDFMLYDALDVLYMDPKLDAFPKLYCCK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGDDHPK 218

RESULT 7

US-09-844-908-6
Sequence 6, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi
Barbosa, Miguel
Li, Gfan

Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908

FILING DATE: 27-Apr-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820

FILING DATE: 12-AUG-1997

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 272 amino acids
TYPE: amino acid

STRANDEDNESS: <unknown>
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-844-908-6

Query Match 100.0%; Score 1179; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWYKIKGLVPTRLLEYLEEKYEHLHYERDEGDKWNNKFEELGLEFPNLPYYID 60
DB 1 MSPILGWYKIKGLVPTRLLEYLEEKYEHLHYERDEGDKWNNKFEELGLEFPNLPYYID 60
QY 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRITAYSKDEFTLV 120
DB 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRITAYSKDEFTLV 120

QY 121 DFLSKLPEMLKAFEDRLCKHTYLNGBDHTVHPDFMLYDALDVLYMDPKLDAFPKLYCCK 180
DB 121 DFLSKLPEMLKAFEDRLCKHTYLNGBDHTVHPDFMLYDALDVLYMDPKLDAFPKLYCCK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGDDHPK 218

RESULT 8

US-09-844-988-3
Sequence 3, Application US/09844988
Patent No. US20020158764A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi
Barbosa, Miguel
Li, Gfan

Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,988

FILING DATE: 26-Apr-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/910,820

FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids
TYPE: amino acid

STRANDEDNESS: <unknown>
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-844-988-3

Query Match 100.0%; Score 1179; DB 9; Length 282;
Best Local Similarity 100.0%; Pred. No. 2.8e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWYKIKGLVPTRLLEYLEEKYEHLHYERDEGDKWNNKFEELGLEFPNLPYYID 60
DB 1 MSPILGWYKIKGLVPTRLLEYLEEKYEHLHYERDEGDKWNNKFEELGLEFPNLPYYID 60
QY 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRITAYSKDEFTLV 120
DB 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRITAYSKDEFTLV 120
QY 121 DFLSKLPEMLKAFEDRLCKHTYLNGBDHTVHPDFMLYDALDVLYMDPKLDAFPKLYCCK 180
DB 121 DFLSKLPEMLKAFEDRLCKHTYLNGBDHTVHPDFMLYDALDVLYMDPKLDAFPKLYCCK 180

QY 181 KRLEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPPK 218
Db 181 KRLEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPPK 218

RESULT 9

US-09-844-988-5

Sequence 5, Application US/09844988

Patent No. US20020158764A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi

Barbosa, Miguel

Li, Gian

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,988

FILING DATE: 26-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/910,820

FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:

NAME: MAKI, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Query Match 100.0%; Score 1179; DB 9; Length 282;

Best Local Similarity 100.0%; Pred. No. 2.8e-111;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKWRKKKTELGLIEFNNLPYYID 60
Db 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKWRKKKTELGLIEFNNLPYYID 60
QY 61 GGVKLTQSAIIRYIADKNNMLGGCKPERAEISMLEGAVLDIRYGSRIAYSDFETLKY 120
Db 61 GGVKLTQSAIIRYIADKNNMLGGCKPERAEISMLEGAVLDIRYGSRIAYSDFETLKY 120
QY 121 DFLSKLPEMLKMFEDRLCKHTYLNQDVTHTPDMFLYDALDVLYMDPMCLDAFPLVCFK 180
Db 121 DFLSKLPEMLKMFEDRLCKHTYLNQDVTHTPDMFLYDALDVLYMDPMCLDAFPLVCFK 180
QY 181 KRLEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPPK 218
Db 181 KRLEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPPK 218

RESULT 10
US-09-844-908-3

Sequence 3, Application US/09844908

Patent No. US20020151021A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi

Barbosa, Miguel

Li, Gian

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,908

FILING DATE: 27-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/910,820

FILING DATE: 12-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: MAKI, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Query Match 100.0%; Score 1179; DB 10; Length 282;

Best Local Similarity 100.0%; Pred. No. 2.8e-111;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKWRKKKTELGLIEFNNLPYYID 60
Db 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKWRKKKTELGLIEFNNLPYYID 60
QY 61 GGVKLTQSAIIRYIADKNNMLGGCKPERAEISMLEGAVLDIRYGSRIAYSDFETLKY 120
Db 61 GGVKLTQSAIIRYIADKNNMLGGCKPERAEISMLEGAVLDIRYGSRIAYSDFETLKY 120
QY 121 DFLSKLPEMLKMFEDRLCKHTYLNQDVTHTPDMFLYDALDVLYMDPMCLDAFPLVCFK 180
Db 121 DFLSKLPEMLKMFEDRLCKHTYLNQDVTHTPDMFLYDALDVLYMDPMCLDAFPLVCFK 180
QY 181 KRLEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPPK 218
Db 181 KRLEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPPK 218

RESULT 11

US-09-844-908-5

Sequence 5, Application US/09844908

```
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
            Zhu, Hengyi
            Barbosa, Miguel
            Li, Gian
            Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-844-908-5

Query Match          100.0%; Score 1179; DB 10; Length 282;
Best Local Similarity 100.0%; Pred. No. 2.8e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIRKIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIRKIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAETSMLEGAVALDIRGVSRIVASKDFETLKV 120
DB 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAETSMLEGAVALDIRGVSRIVASKDFETLKV 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPXCIDAFPKLYCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPXCIDAFPKLYCFK 180
QY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPK 218

RESULT 12
US-09-910-600-26
Sequence 26, Application US/09910600
Publication No. US20030036631A1
GENERAL INFORMATION:
APPLICANT: Longphre, Malinda
APPLICANT: Chang, Han
```

```
APPLICANT: Whitney, Gena
TITLE OF INVENTION: NOVEL SIGLECS AND USES THEREOF
FILE REFERENCE: D0003NP
CURRENT APPLICATION NUMBER: US/09/910,600
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: 60/220,139
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 32
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 26
LENGTH: 298
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: L3cyto-Y641
US-09-910-600-26

Query Match          100.0%; Score 1179; DB 9; Length 298;
Best Local Similarity 100.0%; Pred. No. 3e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIRKIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIRKIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAETSMLEGAVALDIRGVSRIVASKDFETLKV 120
DB 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAETSMLEGAVALDIRGVSRIVASKDFETLKV 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPXCIDAFPKLYCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPXCIDAFPKLYCFK 180
QY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPK 218

RESULT 13
US-10-267-311-25
Sequence 25, Application US/10267311
Publication No. US20030050469A1
GENERAL INFORMATION:
APPLICANT: Stiegel, Marvin
APPLICANT: Chu, N. Randall
APPLICANT: Mizzen, Lee A.
TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
FILE REFERENCE: 12071/002001
CURRENT APPLICATION NUMBER: US/10/267,311
CURRENT FILING DATE: 2002-10-09
PRIOR APPLICATION NUMBER: US/09/613,303
PRIOR FILING DATE: 2000-07-10
PRIOR APPLICATION NUMBER: US 60/143,757
PRIOR FILING DATE: 1999-07-08
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 324
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion sequence
US-10-267-311-25

Query Match          100.0%; Score 1179; DB 9; Length 324;
Best Local Similarity 100.0%; Pred. No. 3.3e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIRKIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIRKIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
```


OY 61 GGVKLTQSMATIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGVSRVIAYSKDFETLKV 120
DB 61 GGVKLTQSMATIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGVSRVIAYSKDFETLKV 120
OY 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
DB 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
OY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDDHPPK 218

RESULT 14
US-09-824-438-6
Sequence 6, Application US/09824438
Publication No. US20030073621A1
GENERAL INFORMATION:
APPLICANT: Butler, Jon P.
APPLICANT: Hale, John E.
APPLICANT: Heath Jr., William F.
APPLICANT: Schoner, Brigitte E.
APPLICANT: Helman, Mark L.
APPLICANT: Becker, Gerald W.
APPLICANT: Varshavsky, Alexander D.
TITLE OF INVENTION: Beta-lipotropin and Uses Thereof
FILE REFERENCE: X-12139
CURRENT APPLICATION NUMBER: US/09/824,438
CURRENT FILING DATE: 2001-10-29
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GST fusion
OTHER INFORMATION: Partner
US-09-824-438-6

Query Match 100.0%; Score 1179; DB 9; Length 331;
Best Local Similarity 100.0%; Pred. No. 3.4e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVOPTRLLLEYLEKEYEHLHYERDEGDKWNNKFFELGLEFPNLPYYTD 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEKEYEHLHYERDEGDKWNNKFFELGLEFPNLPYYTD 60
OY 61 GGVKLTQSMATIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGVSRVIAYSKDFETLKV 120
DB 61 GGVKLTQSMATIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGVSRVIAYSKDFETLKV 120
OY 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
DB 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
OY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDDHPPK 218

RESULT 15
US-09-910-600-22
Sequence 22, Application US/09910600
Publication No. US20030036631A1
GENERAL INFORMATION:
APPLICANT: Longphre, Malinda
APPLICANT: Chang, Han
APPLICANT: Whitney, Gena
TITLE OF INVENTION: NOVEL SILECS AND USES THEREOF
FILE REFERENCE: D0003NP
CURRENT APPLICATION NUMBER: US/09/910,600
CURRENT FILING DATE: 2001-07-20

PRIOR APPLICATION NUMBER: 60/220,139
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 22
LENGTH: 348
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: L3cyto-wt
US-09-910-600-22

Query Match 100.0%; Score 1179; DB 9; Length 348;
Best Local Similarity 100.0%; Pred. No. 3.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVOPTRLLLEYLEKEYEHLHYERDEGDKWNNKFFELGLEFPNLPYYTD 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEKEYEHLHYERDEGDKWNNKFFELGLEFPNLPYYTD 60
OY 61 GGVKLTQSMATIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGVSRVIAYSKDFETLKV 120
DB 61 GGVKLTQSMATIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGVSRVIAYSKDFETLKV 120
OY 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
DB 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
OY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDDHPPK 218

Search completed: May 20, 2003, 12:02:21
Job time : 35.7335 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:09:54 ; Search time 8.41126 Seconds

(without alignments)
2491.578 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179

Sequence: 1 MSPILGWMKIKGLVQPTRL.....IAMPLOGMQATFGGDHPPK 218

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1170	99.2	219	2	A26484 glutathione transt
2	990	84.0	218	2	A45556 glutathione S-tran
3	927	78.6	218	2	A45523 glutathione transt
4	654	55.5	209	2	A48388 glutathione S-tran
5	483.5	41.0	218	2	S33860 glutathione transt
6	483.5	41.0	218	2	A29794 glutathione transt
7	480.5	40.8	218	2	S32425 glutathione transt
8	480.5	40.8	218	2	A47486 glutathione transt
9	478.5	40.6	218	2	S65674 glutathione transt
10	474.5	40.2	218	2	A39375 glutathione transt
11	473.5	40.2	217	2	JX0095 glutathione transt
12	470.5	39.9	218	2	S13202 glutathione transt
13	469.5	39.8	218	2	A23732 glutathione transt
14	468.5	39.7	218	2	B34159 glutathione transt
15	467.5	39.7	218	2	S01719 glutathione transt
16	464.5	39.4	218	2	B29331 glutathione transt
17	464.5	39.4	218	2	A46048 glutathione transt
18	462.5	39.2	218	2	A29036 glutathione transt
19	456.5	38.7	218	1	XURTG4 glutathione transt
20	453.5	38.5	218	2	A46143 mu-class glutathio
21	449.5	38.1	218	2	B28946 glutathione transt
22	449.5	38.1	225	2	A35295 glutathione transt
23	429.5	36.4	230	2	S18464 glutathione transt
24	385.5	32.7	219	2	S50146 major allergen Dpl
25	257.5	21.6	142	2	S17462 glutathione transt
26	248.5	21.1	208	2	S41933 glutathione transt
27	244	20.7	210	2	S71980 glutathione transt
28	242	20.5	203	2	S13780 glutathione transt
29	240	20.4	210	1	A37378 glutathione transt

30	238	20.2	210	2	S71957 glutathione transt
31	237	20.1	210	1	XURTG4 glutathione transt
32	237	20.1	210	1	B55140 glutathione transt
33	236	20.0	210	2	I48112 glutathione S-tran
34	236	20.0	210	2	S71959 glutathione transt
35	235	19.9	210	2	JC6529 glutathione transt
36	233	19.8	209	1	A55140 glutathione transt
37	232	19.7	210	2	A41177 glutathione transt
38	230	19.5	210	2	A49180 glutathione transt
39	219.5	18.6	208	2	S03615 glutathione transt
40	214.5	18.2	223	1	A41031 glutathione transt
41	200.5	17.0	222	2	A26653 glutathione transt
42	198.5	16.8	223	1	A27848 glutathione transt
43	195.5	16.6	222	1	XURTG4 glutathione transt
44	193.5	16.4	221	2	A54858 glutathione transt
45	193.5	16.4	221	2	S24322 glutathione transt

ALIGNMENTS

```

RESULT 1
A26484
glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma japonicum) (fragment)
C:Species: Schistosoma japonicum
C:Date: 25-Oct-1987 #sequence_revision 30-Sep-1989 #text_change 03-Feb-1994
C:Accession: A94139; A26484; A28315
R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 8703-8707, 1986
A:Title: Mr 26,000 antigen of Schistosoma japonicum recognized by resistant MEH1 1:
A:Reference number: A94139; MUID:87041520; PMID:3095841
A:Accession: A94139
A:Molecule type: mRNA
A:Residues: 1-219 <SMID>
R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 6541, 1987
A:Reference number: A94181
A:Contents: annotation; revision to residues 210-219
A:Keywords: glutathione transferase
C:Keywords: transferase

Query Match          99.2%: Score 1170; DB 2; Length 219;
Best Local Similarity 100.0%: Pred. No. 9.6e-95;
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PILDGWMKIKGLVQPTRLLEYLEEKYEHLERDEGDKMRNKKFELGEPNLPYIDGD 62
    |||||||
Db 4 PILDGWMKIKGLVQPTRLLEYLEEKYEHLERDEGDKMRNKKFELGEPNLPYIDGD 63
    |||||||

QY 63 VKLTOSMAIIRYIADKNHMLGCGCPKRAEISMLGCAVIDIRYGVSRAYSDFETLKADF 122
    |||||||
Db 64 VKLTOSMAIIRYIADKNHMLGCGCPKRAEISMLGCAVIDIRYGVSRAYSDFETLKADF 123
    |||||||

QY 123 LSKLPEMKMEEDRLCKHTYVNSGDVHPDMLKDALDVVLYMPCMDAPKIVCFKKR 182
    |||||||
Db 124 LSKLPEMKMEEDRLCKHTYVNSGDVHPDMLKDALDVVLYMPCMDAPKIVCFKKR 183
    |||||||

QY 183 IEAIPQIDKYLKSSKRYIAMPLOGMQATFGGDHPPK 218
    |||||||
Db 184 IEAIPQIDKYLKSSKRYIAMPLOGMQATFGGDHPPK 219
    |||||||

RESULT 2
A45556
glutathione S-transferase - fluke (Schistosoma mansoni)
C:Species: Schistosoma mansoni
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
C:Accession: A45556
R:Wright, M.D.; Harrison, R.A.; Meider, A.M.; Newport, G.R.; Mitchell, G.F.
Mol. Biochem. Parasitol. 49, 177-179, 1991
A:Title: Another 26-kilodalton glutathione S-transferase of Schistosoma mansoni.
A:Reference number: A45556; MUID:92131046; PMID:1775156
A:Accession: A45556

```

A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-218 <NR>
 A>Note: sequence extracted from NCBI backbone (NCBI:77814)
 C:Superfamily: glutathione transferase

Query Match 84.0%; Score 990; DB 2; Length 218;
 Best Local Similarity 82.6%; Pred. No. 4.8e-79;
 Matches 180; Conservative 19; Mismatches 19; Indels 0; Gaps 0;

QY 1 MSPILGKIKGLVQPTLLLEYLEEKYEHLERDEGDKMKKKFELGEPNLPYYID 60
 DB 1 MAPKLGKWKIKGLVQPTLLLEYLEGERYERLYDRNDGVMNKKFKGLDLPNLPYYID 60
 QY 61 GVKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRGVSRATYKDEFTLV 120
 DB 61 GVKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRGVSRATYKDEFTLV 120
 QY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPCLDAFPKLVCFK 180
 DB 121 DFLNQLPGLKMFEDRLSHNTYLNGLDKVTHPDPMLYDALDVLYMDPCLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMATFGGDDHPK 218
 DB 181 ORIEMLPPIKYNLNSRYIKMPLQGSATFGGDAAPK 218

RESULT 3

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <TR>
 A:Cross-references: GB:M21106; NID:g161004; PID:AAA29888.1; PID:g552241
 R:Henkle, K.J.; Davern, K.M.; Wright, M.D.; Ramos, A.J.; Mitchell, G.F.
 M.Ol. Biochem. Parasitol. 40, 23-34, 1990
 A:Title: Comparison of the cloned genes of the 26- and 28-kilodalton glutathione S-trans
 A:Reference number: A44911; MUID:90271935; PMID:1693415
 A:Accession: A45523
 A:Reference number: A45523; MUID:90348716; PMID:2385266
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <TR>
 A:Cross-references: GB:M21106; NID:g161004; PID:AAA29888.1; PID:g552241
 R:Henkle, K.J.; Davern, K.M.; Wright, M.D.; Ramos, A.J.; Mitchell, G.F.
 M.Ol. Biochem. Parasitol. 40, 23-34, 1990
 A:Title: Comparison of the cloned genes of the 26- and 28-kilodalton glutathione S-trans
 A:Reference number: A44911; MUID:90271935; PMID:1693415
 A:Accession: A44941
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 8-218 <HE>
 A:Cross-references: GB:M26914; GB:M26913; NID:g161006; PID:AAA29889.1; PID:g161007
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 78.6%; Score 927; DB 2; Length 218;
 Best Local Similarity 79.8%; Pred. No. 1.5e-73;
 Matches 174; Conservative 15; Mismatches 29; Indels 0; Gaps 0;

QY 1 MSPILGKIKGLVQPTLLLEYLEEKYEHLERDEGDKMKKKFELGEPNLPYYID 60
 DB 1 MAPKLGKWKIKGLVQPTLLLEYLEGERYERLYDRNDGVMNKKFKGLDLPNLPYYID 60
 QY 61 GVKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRGVSRATYKDEFTLV 120
 DB 61 GVKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRGVSRATYKDEFTLV 120
 QY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPCLDAFPKLVCFK 180
 DB 121 DFLNQLPGLKMFEDRLSHNTYLNGLDKVTHPDPMLYDALDVLYMDPCLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMATFGGDDHPK 218
 DB 181 KCIEDLPQIKYNLNSRYIKMPLQGSATFGGDDHPK 218

RESULT 4

A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-209 <NR>
 A:Experimental source: Puerto Rico
 A:Note: sequence analysis of a Fasciola hepatica glutathione S-transferase cDNA clon
 A:Superfamily: glutathione transferase

Query Match 55.5%; Score 654; DB 2; Length 209;
 Best Local Similarity 58.2%; Pred. No. 9.6e-50;
 Matches 121; Conservative 31; Mismatches 56; Indels 0; Gaps 0;

QY 10 IKGVLQPTLLLEYLEEKYEHLERDEGDKMKKKFELGEPNLPYYIDGDKLTQSM 69
 DB 1 IRGLQPVRLLEYLEGERYERLYDRNDGVMNKKFKGLDLPNLPYYIDCKLTQSV 60
 QY 70 AIIRYIADKHNMLGGCPKRAEISMLEGAVLDIRGVSRATYKDEFTLVDFLSKLPK 129
 DB 61 AIIRYIADKHNMLGGCPKRAEISMLEGAVLDIRGVSRATYKDEFTLVDFLSKLPK 120
 QY 130 LKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPCLDAFPKLVCFKRIEAIPOI 189
 DB 121 LKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPCLDAFPKLVCFKRIEAIPOI 180
 QY 190 DYLKSSKIYAMPLOGMATFGGDDHP 217
 DB 181 KAMSESEKIKMPLNSWMSFSGGDAAP 208

RESULT 5

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-218 <REI>
 A:Cross-references: GB:U13448
 A:Note: the authors translated the codon GAG for residue 49 as Gly and GAG for residu
 R:Townsend, A.J.; Goldsmith, M.E.; Pickett, C.B.; Cowan, K.H.
 J. Biol. Chem. 264, 21582-21590, 1989
 A:Title: Isolation, characterization, and expression in Escherichia coli of two murin
 A:Reference number: A34159; MUID:90094327; PMID:2689439
 A:Accession: A34159
 A:Molecule type: mRNA
 A:Residues: 1-218 <TON>
 A:Cross-references: GB:U04632; NID:g193547; PID:AAA37705.1; PID:g309260
 R:Pearson, W.R.; Reinhardt, J.; Sisk, S.C.; Anderson, K.S.; Adler, P.N.
 J. Biol. Chem. 263, 13324-13332, 1988
 A:Title: Tissue-specific induction of murine glutathione transferase mRNAs by butylat
 A:Reference number: A92668; MUID:88330838; PMID:3411659
 A:Contents: Clone PGT 875
 A:Accession: A28946
 A:Molecule type: mRNA
 A:Residues: 2-218 <PEA>

A:Cross-references: GB:J03952; NID:9193687; PIDN:AAA37747.1; PID:9309278
 R:Pearson, W.R.; Windle, J.J.; Morrow, J.F.; Benson, A.M.; Talalay, P.
 J. Biol. Chem. 258, 2052-2062, 1983
 A:Title: Increased synthesis of glutathione S-transferases in response to anticarcinogen
 A:Reference number: A92411; MUID:83109018; PMID:6822548
 A:Accession: A20831
 A:Molecule type: protein
 A:Residues: 2-7, 'X', 8-13, 'X', 15-23, 'X', 25, 'XX', 28, 'N', 30-31, 'X', 33, 'X', 35-39, 'W', 41 <PEI
 R:Mannervik, B.; Alin, P.; Guttenberg, C.; Jansson, H.; Tahr, M.K.; Warholm, M.; Jorvæ
 Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
 A:Title: Identification of three classes of cytosolic glutathione transferase common to
 A:Reference number: A24735; MUID:86042634; PMID:3864155
 A:Accession: I24735
 A:Molecule type: protein
 A:Residues: 2-7, 'X', 8-13, 'X', 15-23, 'X', 25 <MAN>
 R:Hayes, J.D.; Kerr, L.A.; Peacock, S.D.; Cronshaw, A.D.; Mclellan, L.I.
 Biochem. J. 277, 501-512, 1991
 A:Title: Hepatic glutathione S-transferases in mice fed on a diet containing the anticar
 adent elution of the glutathione-S-transferase affinity matrix.
 A:Reference number: S16933; MUID:91315425; PMID:1859377
 A:Accession: S16933
 A:Molecule type: protein
 A:Residues: 2-36 <NAV>
 R:Fernandes, C.L.; Dong, J.H.; Roebuck, B.D.; Chisari, F.V.; Montali, J.A.; Schmidt Jr.,
 Arch. Biochem. Biophys. 331, 104-116, 1996
 A:Title: Elevations of hepatic quinone reductase, glutathione, and alpha- and mu-class g
 A:Reference number: S71311; MUID:96268466; PMID:8660689
 A:Accession: S71312
 A:Molecule type: protein
 A:Residues: 97-108 <FEK>
 C:Genetics:
 A:Gene: GSTM1
 A:Intons: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase
 F:2-218/Product: glutathione transferase mul #status experimental <MAT>
 Query Match 41.0%; Score 483.5; DB 2; Length 218;
 Best Local Similarity 44.2%; Pred. No. 7.6e-35;
 Matches 92; Conservative 38; Mismatches 73; Indels 5; Gaps 1;
 QY 1 MSPILGYWKIKGVOPTRLLLEYLEEYEEHLERDEG-----DKMNKKEFELGLEPRL 55
 DB 1 MPMLIGWVNRGLTHPRLLLEYLEEYEEHLERDEG-----DKMNKKEFELGLEPRL 60
 QY 56 PYIIDGVKLTQSMATIRYADKNMIGCCPKERAELISMEGAVLDIRYGVSRIVASKDF 115
 DB 61 PYIIDSHKRTQSNALIRYARKNHHDGTEEEERIRADIVENQVMOTRMOLIMCTVDPF 120
 QY 116 ETLLKVPFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVVLYMDPMLDAEPK 175
 DB 121 EKQKPEFLKTIPEKMKIYSEFLGKRPFADGKYTVDFLAYDILDQYRMEPCIDAFPN 180
 QY 176 LVCFKKRIEAIPOIDKYKSKYIAMP 203
 DB 181 LRDFLAFEGKLKRIISAYMKSSRYIATPI 208
 RESULT 6
 A29794
 glutathione transferase (EC 2.5.1.18) class mu chain 3 - rat
 N:Alternate names: glutathione S-transferase Yb1; ligandin
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 28-Aug-1989 #sequence, revision 28-Aug-1989 #text, change 18-Jun-1999
 R:Accession: A29794; A25510; A24085; B61004; E24735; A33397; S27111; S17167; A26187
 R:Chang, C.; Saltzman, A.G.; Sorensen, N.S.; Hlipkka, R.A.; Liao, S.
 J. Biol. Chem. 262, 11901-11903, 1987
 A:Title: Identification of glutathione S-transferase Yb-1 mRNA as the androgen-repressed
 A:Reference number: A29794; MUID:87308179; PMID:3040722
 A:Accession: A29794
 A:Molecule type: mRNA
 A:Residues: 1-218 <CHA>
 A:Cross-references: GB:J02810; NID:9204514; PIDN:AAA1293.1; PID:9204515

R:Lat, H.C.T.; Grove, G.; Tu, C.P.D.
 Nucleic Acids Res. 14, 6101-6114, 1986
 A:Title: Cloning and sequence analysis of a cDNA for a rat liver glutathione S-tran
 A:Reference number: A25510; MUID:86312882; PMID:2875437
 A:Accession: A25510
 A:Molecule type: mRNA
 A:Residues: 1-218 <LAT>
 A:Cross-references: GB:X04229; NID:956337; PIDN:CAA27811.1; PID:956338
 R:Ding, G.J.F.; Lu, A.Y.H.; Pickett, C.B.
 J. Biol. Chem. 260, 13268-13271, 1985
 A:Reference number: A24085; MUID:86033768; PMID:3840477
 A:Accession: A24085
 A:Molecule type: mRNA
 A:Residues: 1-198, 'NC', 201-218 <DIN>
 A:Cross-references: GB:M11719; NID:9204502; PIDN:AAA1287.1; PID:9204503
 A:Experimental source: clone pGTR/C44
 R:Chang, L.H.; Hsieh, J.C.; Chen, W.L.; Tam, M.F.
 Electrophoresis 11, 589-593, 1990
 A:Title: Identification of rat liver glutathione S-transferase Yb subunits by parti
 ical isoelectric focusing gel.
 A:Reference number: A61004; MUID:91031411; PMID:2226415
 A:Accession: B61004
 A:Molecule type: protein
 A:Residues: 2-26 <CH2>
 R:Mannervik, B.; Alin, P.; Guttenberg, C.; Jansson, H.; Tahr, M.K.; Warholm, M.; J
 Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
 A:Title: Identification of three classes of cytosolic glutathione transferase commc
 A:Reference number: A24735; MUID:86042634; PMID:3864155
 A:Accession: F24735
 A:Molecule type: protein
 A:Residues: 2-20 <MAN>
 R:Hsieh, J.C.; Liu, L.F.; Chen, W.L.; Tam, M.F.
 Biochem. Biophys. Res. Commun. 162, 1147-1154, 1989
 A:Title: Expression of Yb-1 glutathione S-transferase using a baculovirus expressic
 A:Reference number: A33397; MUID:89350924; PMID:2669745
 A:Accession: A33397
 A:Molecule type: protein
 A:Residues: 2-21, 212-218 <HS1>
 R:Katsumi, R.M.; Bono, B.; Colman, R.F.
 Arch. Biochem. Biophys. 298, 667-677, 1992
 A:Title: Identification of Tyr(115) labeled by S-(4-Dromo-2,3-dioxobutyl)glutathior
 A:Reference number: S27111; MUID:93037509; PMID:1416995
 A:Accession: S27111
 A:Molecule type: protein
 A:Residues: 2-20, 83-86, 'X', 88-96; 109-115, 'X', 117-122 <KAT>
 R:Hsieh, J.C.; Huang, S.C.; Chen, W.L.; Lai, Y.C.; Tam, M.F.
 Biochem. J. 278, 293-297, 1991
 A:Title: Cysteine-86 is not needed for the enzymic activity of glutathione S-transf
 A:Reference number: S17167; MUID:91354218; PMID:1883338
 A:Accession: S17167
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 2-218 <HS2>
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase
 Query Match 41.0%; Score 483.5; DB 2; Length 218;
 Best Local Similarity 44.2%; Pred. No. 7.6e-35;
 Matches 92; Conservative 38; Mismatches 73; Indels 5; Gaps 1;
 QY 1 MSPILGYWKIKGVOPTRLLLEYLEEYEEHLERDEG-----DKMNKKEFELGLEPRL 55
 DB 1 MPMLIGWVNRGLTHPRLLLEYLEEYEEHLERDEG-----DKMNKKEFELGLEPRL 60
 QY 56 PYIIDGVKLTQSMATIRYADKNMIGCCPKERAELISMEGAVLDIRYGVSRIVASKDF 115
 DB 61 PYIIDSHKRTQSNALIRYARKNHHDGTEEEERIRADIVENQVMOTRMOLIMCTVDPF 120
 QY 116 ETLLKVPFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVVLYMDPMLDAEPK 175
 DB 121 EKQKPEFLKTIPEKMKIYSEFLGKRPFADGKYTVDFLAYDILDQYRMEPCIDAFPN 180
 QY 176 LVCFKKRIEAIPOIDKYKSKYIAMP 203

Db 181 LKDFLARFEGLEKISAYMKSSRFLPTRYMAVWG 208

RESULT 7

S32425

glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 2) - human

N:Alternate names: glutathione transferase mu4

C:Species: Homo sapiens (man)

C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 18-Jun-1999

C:Accession: S32425; S29337

R:Zhong, S.; Spurr, N.K.; Hayes, J.D.; Wolf, C.R.

Biochem. J. 291, 41-50, 1993

A:Title: Deduced amino acid sequence, gene structure and chromosomal location of a novel

A:Reference number: S32424; MUID:93228631; PMID:8471052

A:Accession: S32425

A:Molecule type: DNA

A:Residues: 1-218 <2HO>

A:Cross-references: EMBL:X68677; NID:g31934; PIDN:CAA48637.1; PID:g31935

A:Note: the authors translated the codon AAG for residue 182 as Arg and CCA for residue

C:Genetics:

A:Gene: GDB:GSTM4

A:Cross-references: GDB:134191; OMIM:138333

A:Map position: 1p13.3-1p13.3

A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3

C:Superfamily: glutathione transferase

C:Keywords: dimer; transferase

Query Match

Best Local Similarity 40.8%; Score 480.5; DB 2; Length 218;

Matches 95; Conservative 37; Mismatches 79; Indels 5; Gaps 1;

1 MSPILGVRKINGVOPRLLEYLEEKEEHLRYERD-----EGDKMRKKELEPPNL 55

1 MPMLGVRKINGVOPRLLEYLEEKEEHLRYERD-----EGDKMRKKELEPPNL 60

56 PYIDGVKRLTOSMAIRIYADKNHMLGGCPKERAISMLEGAVLDIRYGSRIAYSKDF 115

61 PYLDGAKKITQSNAILCYIAKRNHMLGCEEEKIRVDILENQAMDVSNQALRCYSPDF 120

116 ETLKVDPLSKPEMLKMFEDRLCHKTYLNGDVTHTPDMFLYDALDYLYMDPMLCLAPPK 175

121 EKLKPEYLEELPTMQHRSQFLGKRPWFVGDKIFVDFLAYVDLHRIFFPNCIDAPFN 180

176 LVCFKRRLEAIPQIDKYLKSSKITYAMPLOGQATFG 211

181 LKDFISRFEGLEKISAYMKSSRFLPKPLTRYMAVWG 216

Db

RESULT 8

A47486

glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 1) - human

C:Species: Homo sapiens (man)

C>Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 18-Jun-1999

C:Accession: A47486; B47486; S36782; I37438; S45685

R:Johnson, K.E.; Johnson, K.J.; Rifenbery, D.; Henner, W.D.

J. Biol. Chem. 268, 16958-16965, 1993

A:Title: Isolation and analysis of the gene and cDNA for a human mu class glutathione S-

A:Reference number: A47486; MUID:93352467; PMID:8349586

A:Accession: A47486

A:Molecule type: DNA

A:Residues: 1-218 <COM>

A:Cross-references: GB:M96233

A:Accession: B47486

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-218 <C2>

A:Cross-references: GB:M96234; NID:g306818; PIDN:AAA57347.1; PID:g306819

A:Experimental source: Hela cells

R:Rosa, V.L.; Board, P.G.

Biochem. J. 294, 373-380, 1993

A:Title: Molecular cloning and heterologous expression of an alternatively spliced human

A:Reference number: S36782; MUID:93384505; PMID:8373352

A:Accession: S36782

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-218 <ROS>

A:Cross-references: EMBL:M99422

A:Experimental source: testis

R:Comstock, K.E.; Wadsten, M.; Hao, X.Y.; Henner, W.D.; Mannervik, B.

Arch. Biochem. Biophys. 311, 487-495, 1994

A:Title: A comparison of the enzymatic and physicochemical properties of human glutat

A:Reference number: S45684; MUID:94263230; PMID:8203914

A:Contents: annotation

R:Taylor, J.B.; Oliver, J.; Sherrington, R.; Pemble, S.E.

Biochem. J. 274, 587-593, 1991

A:Title: Structure of human glutathione S-transferase class Mu genes.

A:Reference number: I37438; MUID:91174774; PMID:2006920

A:Accession: I37438

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 39-120 <RES>

A:Cross-references: EMBL:X68637; NID:g31936; PIDN:CAA0167.1; PID:g31937

C:Genetics:

A:Gene: GDB:GSTM4

A:Cross-references: GDB:134191; OMIM:138333

A:Map position: 1p13.3-1p13.3

A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3

C:Superfamily: glutathione transferase

C:Keywords: alternative splicing; dimer; transferase

Query Match

Best Local Similarity 40.8%; Score 480.5; DB 2; Length 218;

Matches 95; Conservative 37; Mismatches 79; Indels 5; Gaps 1;

1 MSPILGVRKINGVOPRLLEYLEEKEEHLRYERD-----EGDKMRKKELEPPNL 55

1 MPMLGVRKINGVOPRLLEYLEEKEEHLRYERD-----EGDKMRKKELEPPNL 60

56 PYIDGVKRLTOSMAIRIYADKNHMLGGCPKERAISMLEGAVLDIRYGSRIAYSKDF 115

61 PYLDGAKKITQSNAILCYIAKRNHMLGCEEEKIRVDILENQAMDVSNQALRCYSPDF 120

116 ETLKVDPLSKPEMLKMFEDRLCHKTYLNGDVTHTPDMFLYDALDYLYMDPMLCLAPPK 175

121 EKLKPEYLEELPTMQHRSQFLGKRPWFVGDKIFVDFLAYVDLHRIFFPNCIDAPFN 180

176 LVCFKRRLEAIPQIDKYLKSSKITYAMPLOGQATFG 211

181 LKDFISRFEGLEKISAYMKSSRFLPKPLTRYMAVWG 216

Db

RESULT 9

S65674

glutathione transferase (EC 2.5.1.18) class mu chain 7.8 - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Jun-1999

C:Accession: S65674; S30380

R:Lee, S.H.; Lee, Y.S.; Han, J.S.; Kim, Y.S.; Koh, J.K.

Arch. Biochem. Biophys. 318, 424-429, 1995

A:Title: Cloning and expression of a cDNA for Mu-class Glutathione S-transferase from

A:Reference number: S65674; MUID:95251394; PMID:7733673

A:Accession: S65674

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-218 <LEP>

A:Cross-references: EMBL:L23766; NID:g388198; PIDN:AAA69665.1; PID:g388199

R:Primiano, T.; Novak, R.F.

Arch. Biochem. Biophys. 301, 404-410, 1993

A:Title: Purification and characterization of class mu glutathione S-transferase iso2

A:Reference number: S30380; MUID:93213177; PMID:8460949

A:Accession: S30380

A:Molecule type: Protein

A:Residues: 2-21 <PRI>

C:Superfamily: glutathione transferase

C:Accession: S13202
 R:Robert de Saint Vincent, B.; Hyrien, O.; Debatisse, M.; Buttin, G.
 Eur. J. Biochem. 193; 19-24, 1990
 A:Title: Clonification of mu class glutathione S-transferase genes and an adenylate de
 A:Reference number: S13202; MUID:91031445; PMID:2226437
 A:Accession: S13202
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <ROB>
 A:Cross-references: GB:X57489; NID:949541; PIDN:CAA40726.1; PID:949542
 A:Note: In Genbank entry CLV1, release 109, the source is designated as Cricetus longi
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 39.9%; Score 470.5; DB 2; Length 218;
 Best Local Similarity 43.3%; Pred. No. 1e-33;
 Matches 90; Conservative 36; Mismatches 77; Indels 5; Gaps 1;

1 MSPILGKIKGLVPTLLLEKEEHLERDEG----DKWRNKKFELGEPNL 55
 1 MPMLGKNNVGLNPILLETYSSEKKYMGDAPDSRQSWLNEKFKLGDPVL 60
 56 PYIDGVKLTQSMALIRYADKNNMLGCPKRAEISMLEGAVDIRGVSRVSKDF 115
 61 PYLDGSHKITQSNAILRYARKNMLCGETEERIRVDIVENQADTRMQLMLCYNPDF 120
 116 ETLKVDPLSKLPEMLKMFEDRLCHKTLYNGDVHPDMLYDALDVLVYMDPMCIDAPPK 175
 121 EKQREFLKTPEKKMKMSEFLGKOPMFGADKVTLCGFLAVDLVQYMFEPKCLDPPN 180
 176 LVCFKKRIEAIPOIDKYLKSSKYIAPL 203
 181 LKDFLAREGLKKSISAVYKTSRFLRPI 208

RESULT 13

A23732
 glutathione transferase (EC 2.5.1.18) mu - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 18-Jun-1999
 C:Accession: A23732; S21908
 R:Norris, J.S.; Schwartz, D.A.; Macleod, S.L.; Fan, W.; O'Brien, T.J.; Harris, S.E.; Tr
 Mol. Endocrinol. 5, 979-986, 1991
 A:Title: Cloning of a mu-class glutathione S-transferase complementary DNA and character
 A:Reference number: A23732; MUID:92049380; PMID:1944302
 A:Accession: A23732
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <NOR>
 A:Cross-references: EMBL:X61033; NID:949638; PIDN:CAA43368.1; PID:949639
 A:Note: submitted to the EMBL Data Library, July 1991
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 39.8%; Score 469.5; DB 2; Length 218;
 Best Local Similarity 44.3%; Pred. No. 1.3e-33;
 Matches 90; Conservative 37; Mismatches 71; Indels 5; Gaps 1;

1 MSPILGKIKGLVPTLLLEKEEHLERDEG----DKWRNKKFELGEPNL 55
 1 MPVLGYMDIRGLAHAILLETYSSEKKYMGDAPNFDPSQWLNKFKLGIDPVL 60
 56 PYIDGVKLTQSMALIRYADKNNMLGCPKRAEISMLEGAVDIRGVSRVSKDF 115
 61 PYLDGSHKITQSNAILRYARKNMLCGETEERIRVDIVENQADTRMQLMLCYNPDF 120
 116 ETLKVDPLSKLPEMLKMFEDRLCHKTLYNGDVHPDMLYDALDVLVYMDPMCIDAPPK 175
 121 EKQREFLKTPEKKMKMSEFLGKOPMFGADKVTLCGFLAVDLVQYMFEPKCLDPPN 180
 176 LVCFKKRIEAIPOIDKYLKSSKY 198
 181 LKDFLAREGLKKSISAVYKTSRFLRPI 203

RESULT 14

B34159
 glutathione transferase (EC 2.5.1.18) mu2 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Apr-1990 #sequence_revision 25-Sep-1992 #text_change 18-Jun-1999
 C:Accession: B34159; S30369; S30370; S30372
 R:Townsend, A.J.; Goldsmith, M.E.; Pickett, C.B.; Cowan, K.H.
 J. Biol. Chem. 264, 21582-21590, 1989
 A:Title: Isolation, characterization, and expression in Escherichia coli of two murin
 A:Reference number: A34159; MUID:90094327; PMID:2689439
 A:Accession: B34159
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <TOW>
 A:Cross-references: GB:J04696; NID:9193549; PIDN:AAA37706.1; PID:9309261
 Arch. Biochem. Biophys. 301, 143-150, 1993
 A:Title: Purification and characterization of glutathione S-transferase of murine ova
 A:Reference number: S30369; MUID:93183007; PMID:8442656
 A:Accession: S30369
 A:Molecule type: protein
 A:Residues: 2-25 <AM1>
 A:Note: this form was identified by pi 6.7
 A:Accession: S30370
 A:Molecule type: protein
 A:Residues: 2-25 <AM2>
 A:Note: this form was identified by pi 7.9
 A:Accession: S30372
 A:Molecule type: protein
 A:Residues: 2-25 <AM3>
 A:Note: this form was identified by pi 8.5
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 39.7%; Score 468.5; DB 2; Length 218;
 Best Local Similarity 42.8%; Pred. No. 1.6e-33;
 Matches 89; Conservative 41; Mismatches 73; Indels 5; Gaps 1;

1 MSPILGKIKGLVPTLLLEKEEHLERDEG----DKWRNKKFELGEPNL 55
 1 MPMTLYGMDIRGLAHAILLETYSSEKKYMGDAPDPSRQSWLSEKFKLGIDPVL 60
 56 PYIDGVKLTQSMALIRYADKNNMLGCPKRAEISMLEGAVDIRGVSRVSKDF 115
 61 PYLDGSHKITQSNAILRYARKNMLCGETEERIRVDIVENQADTRMQLMLCYNPDF 120
 116 ETLKVDPLSKLPEMLKMFEDRLCHKTLYNGDVHPDMLYDALDVLVYMDPMCIDAPPK 175
 121 EKQREFLKTPEKKMKMSEFLGKOPMFGADKVTLCGFLAVDLVQYMFEPKCLDPPN 180
 176 LVCFKKRIEAIPOIDKYLKSSKYIAPL 203
 181 LKDFMGREGGLKKSIDYKSSRFLSKPI 208

RESULT 15

S01719
 glutathione transferase (EC 2.5.1.18) class mu, GSTW1 - human
 N:Alternate names: glutathione S-transferase 1, mu (H-b); glutathione transferase 4;
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 21-Jul-2000
 C:Accession: S01719; A30770; S32424; S13905; B22457; D35187; S27188; I51867;
 R:DeLong, J.L.; Chang, C.M.; Wang-Peng, J.; Knutsen, T.; Tu, C.P.D.
 Nucleic Acids Res. 16, 8541-8554, 1988
 A:Title: The human liver glutathione S-transferase gene superfamily: expression and c
 A:Reference number: S01719; MUID:883335606; PMID:3419925
 A:Contents: allele 1a (mu)
 A:Accession: S01719
 A:Molecule type: mRNA
 A:Residues: 1-218 <DEJ>
 A:Cross-references: EMBL:X08020; NID:931923; PIDN:CAA30821.1; PID:931924

A:Experimental source: liver
 R:Seidegard, J.; Vorachek, W.R.; Pero, R.W.; Pearson, W.R.
 Proc. Natl. Acad. Sci. U.S.A. 85, 7293-7297, 1988
 A:Title: Hereditary differences in the expression of the human glutathione transferase A
 A:Reference number: A30770; MUID:89017184; PMID:3174634
 A:Contents: allele 1b (psi)
 A:Accession: A30770
 A:Molecule type: mRNA
 A:Residues: 1-172, 'N', 174-218 <SEI>
 A:Cross-references: EMBL:003617; NID:9183668; PIDN:AAA59203.1; PID:g306812
 A:Experimental source: liver
 R:Zhong, S.; Spurr, N.K.; Hayes, J.D.; Wolf, C.R.
 Biochem. J. 291, 41-50, 1993
 A:Title: Deduced amino acid sequence, gene structure and chromosomal location of a novel
 A:Reference number: S32424; MUID:93228631; PMID:8471052
 A:Accession: S32424
 A:Molecule type: DNA
 A:Residues: 1-43, 'T', 45-172, 'N', 174-218 <ZHO>
 A:Cross-references: EMBL:X68676
 A:Note: the authors translated the codon ACG for residue 44 as Ser
 R:Comstock, K.E.; Sanderson, B.J.; Clafilin, G.; Henner, W.D.
 Nucleic Acids Res. 18, 3670, 1990
 A:Title: GST1 gene deletion determined by polymerase chain reaction.
 A:Reference number: I37437; MUID:90301515; PMID:2362832
 A:Accession: I37437
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME
 A:Molecule type: DNA
 A:Residues: 60-118 <RES>
 A:Cross-references: EMBL:X51451; NID:g31922; PIDN:CAA35617.1; PID:g4378985
 R:Singhal, S.S.; Ahmad, H.; Sharma, R.; Gupta, S.; Haque, A.K.; Awasthi, Y.C.
 Arch. Biochem. Biophys. 285, 64-73, 1991
 A:Title: Purification and characterization of human muscle glutathione S-transferases: E
 A:Reference number: S13905; MUID:91119426; PMID:1846734
 A:Accession: S13905
 A:Molecule type: protein
 A:Residues: 2-15 <SIN>
 A:Experimental source: muscle
 R:Allin, P.; Mannervik, B.; Jornvall, H.
 FEBS Lett. 182, 319-322, 1985
 A:Title: Structural evidence for three different types of glutathione transferase in hum
 A:Reference number: A91336; MUID:8515454; PMID:397955
 A:Accession: B22457
 A:Molecule type: protein
 A:Residues: 2-24 <ALI>
 R:Tsuchida, S.; Makl, T.; Sato, K.
 J. Biol. Chem. 265, 7150-7157, 1990
 A:Title: Purification and characterization of glutathione transferases with an activity
 A:Reference number: A35187; MUID:90237002; PMID:2110160
 A:Accession: D35187
 A:Molecule type: protein
 A:Residues: 2-10, 'X', 12-14, 'X', 16-17, 'X', 19-25 <TSU>
 A:Experimental source: heart
 R:Singhal, S.S.; Saxena, M.; Awasthi, S.; Ahmad, H.; Sharma, R.; Awasthi, Y.C.
 Biochem. Biophys. Acta 1171, 19-26, 1992
 A:Title: Gender related differences in the expression and characteristics of glutathione
 A:Reference number: S27188; MUID:93042004; PMID:1420361
 A:Accession: S27188
 A:Molecule type: protein
 A:Residues: 2-13 <S12>
 A:Experimental source: colon
 R:Pearson, W.R.; Vorachek, W.R.; Xu, S.J.; Berger, R.; Hart, I.; Vannals, D.; Patterson,
 Am. J. Hum. Genet. 53, 220-233, 1993
 A:Title: Identification of class-mu glutathione transferase genes GSTM1-GSTM5 on human C
 A:Reference number: I51867; MUID:93304417; PMID:8317488
 A:Accession: I51867
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 125-172, 'N', 174-186 <RE2>
 A:Cross-references: GB:S62935; NID:g386429; PIDN:AAD13938.1; PID:g4261638
 C:Comment: The GSTM1 locus for the mu isoenzyme of glutathione transferase is polymorphi
 stillbene ozide.
 C:Genetics:
 A:Gene: GDB:GSTM1; GSTM16

A:Cross-references: GDB:120020; OMIM:138350
 A:Map position: 1p13.3-1p13.3
 A:introns: 12/3; 36/1; 59/3; 87/1; 120/3; 152/3
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; polymorphism; transferase
 F:2-218/Product: glutathione transferase class mu, GSTM1 #status predicted <Mat>
 Query Match 39.7%; Score 467.5; DB 2; Length 218;
 Best Local Similarity 42.6%; Pred. No. 1.9e-33;
 Matches 92; Conservative 39; Mismatches 80; Indels 5; Gaps 1;
 QY 1 MSPILGYMKIKGLVPTRLLEYLEEYEEHLETERDEG-----DKRNKKFELGLEFPNU 55
 DB 1 MPAILGYMKIRGLAARILLLETDSEYEKKTTPDADPDYDSQMLNEKFKGLDFPN 60
 QY 56 PYIIDGVKLTOSMAIRYIADKNNMLGGCPKRAEISMLEGAVLDIRYGVSRVAYSXDF 115
 DB 61 PYLIDGAHKITOSNALICYIARKNNLGGEEFEKRIYDLENGTMDNHHQMGICVNPPE 120
 QY 116 ETLKVPFLSKLPMLKMFEDRLCHKTYTLNGDHTHDFMLYDALDYVLYMDPGLDAFR 175
 DB 121 EKLKPYLEELPEKLYSEFLGRPWFAGNKITFDVFLYDVLDRIFEPKCLDAFPN 180
 QY 176 IVCFKRIPAIPOIDRYLKSSKRYIAMPLOGQATFG 211
 DB 181 LKDFISRFEGLEKISAYMKSSRFLPRPVSMAVWG 216

Search completed: May 20, 2003, 11:39:23
 Job time : 9.41126 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 5.14571 Seconds

(without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179
Sequence: 1 MSPILGYWKIKGLVQPTRL.....IAMPLOGMQLTGGGDHPK 218

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1179	100.0	218	1	GR26_SCHUA
2	990	84.0	218	1	GR27_SCHUA
3	927	78.6	218	1	GR26_SCHUA
4	709.5	60.2	220	1	GR29_FASHE
5	699	59.3	217	1	GR27_FASHE
6	681	57.8	217	1	GR28_FASHE
7	680	57.7	217	1	GR26_FASHE
8	481.5	40.8	217	1	GRM1_MOUSE
9	481.5	40.8	217	1	GRM1_RAT
10	480.5	40.8	218	1	GRM4_HUMAN
11	478.5	40.6	218	1	GRM6_MOUSE
12	477.5	40.5	217	1	GRM6_RABIT
13	473.5	40.2	217	1	GRM2_HUMAN
14	473.5	40.2	217	1	GRM2_CAVPO
15	468.5	38.7	217	1	GRM2_CRILLO
16	468.5	39.7	217	1	GRM2_MESAU
17	467.5	39.7	217	1	GRM2_MOUSE
18	465.5	39.5	217	1	GRM1_HUMAN
19	463.5	39.3	217	1	GRM5_HUMAN
20	461.5	39.1	217	1	GRM2_RAT
21	455.5	38.6	217	1	GRM2_RAT
22	452.5	38.4	217	1	GRM3_MOUSE
23	448.5	38.0	224	1	GRM3_HUMAN
24	437.5	37.1	224	1	GRM3_MOUSE
25	432.5	36.7	219	1	GRM2_CHICK
26	385.5	32.7	219	1	GRM1_DERPT
27	251.5	21.3	208	1	GRP_DIRIM
28	248.5	21.1	208	1	GRP_ONCVO
29	246	20.9	210	1	GRP1_BUFBU
30	244	20.7	209	1	GRP_MESAU
31	242	20.5	207	1	GRP_PIG
32	240	20.4	209	1	GRP_HUMAN
33	238	20.2	209	1	GRP_CRIMI

34	237	20.1	209	1	GRP2_MOUSE	P19157 mus musculus
35	237	20.1	209	1	GRP_RAT	P04906 ratius norv
36	236	20.0	209	1	GRP_CRILLO	P46424 cricetus
37	235	19.9	209	1	GRP_MACMU	O28514 macaca mula
38	233	19.8	209	1	GRP1_MOUSE	P46425 mus musculus
39	230	19.5	209	1	GRP_BOVIN	P28801 bos taurus
40	219.5	18.6	208	1	GRP_CAEEL	P10299 caenorhabdi
41	214.5	18.2	223	1	GRM1_RABIT	O08863 oryctolagus
42	204.5	17.3	210	1	GRP2_BUFBU	P83325 bufo bufo (
43	200.5	17.0	221	1	GRM2_RAT	P04903 ratius norv
44	198.5	16.8	221	1	GRM2_MOUSE	P10648 mus musculus
45	198.5	16.8	222	1	GRM1_MOUSE	P13745 mus musculus

ALIGNMENTS

RESULT 1
GR26_SCHUA STANDARD: PRT; 218 AA.
ID GR26_SCHUA
AC P08515;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (Gst 26) (S26 antigen)
DE (Gst class-alpha).
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
[1]
RP SEQUENCE FROM N.A.
RP MEDLINE=87041520; PubMed=3095841;
RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,
RA Mitchell G.F.;
RT "Mr 26,000 antigen of Schistosoma japonicum recognized by resistant
RT WEHI 129/J mice is a parasite glutathione S-transferase.",
RL Proc. Natl. Acad. Sci. U.S.A. 83:8703-8707(1986).
RN [2]
RP REVISIONS.
RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,
RA Mitchell G.F.;
RL Proc. Natl. Acad. Sci. U.S.A. 84:6541-6541(1987).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=95276631; PubMed=7538846;
RA Lim K., Ho J.X., Keeling K., Gilliland G.L., Ji X., Ruker F.,
RA Carter D.C.;
RT "Three-dimensional structure of Schistosoma japonicum glutathione S-
RT transferase fused with a six-amino acid conserved neutralizing
RT epitope of gp41 from HIV.",
RL Protein Sci. 3:2233-2244(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95156484; PubMed=7853399;
RA McTigue M.A., Williams D.R., Tainer J.A.;
RT "Crystal structures of a schistosomal drug and vaccine target:
RT glutathione S-transferase from Schistosoma japonica and its complex
RT with the leading antischistosomal drug praziquantel.",
RL J. Mol. Biol. 246:21-27(1995).
CC -I- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -I- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -I- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -I- SUBUNIT: HOMODIMER.
CC -I- MISCELLANEOUS: THERE ARE AT LEAST TWO ISOENZYMES OF GST IN
CC S.JAPONICUM.
CC -I- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: M14654; AAB59203.1; -
DR PIR: A26484; A26484.
DR PDB: 1GNE: 30-NOV-94.
DR PDB: 1GTA: 07-FEB-95.
DR PDB: 1GTB: 01-DEC-95.
DR PDB: 1B8X: 12-APR-99.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C: 1.
DR Pfam: PF02798; GST_N: 1.
DR Transferrase; Antigen; Multigene family; 3D-structure.
SQ SEQUENCE 218 AA: 25498 MW: 562AC418BD0EF13F CRC64;

Query Match 100.0%; Score 1179; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.8e-98;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTQSAAIRYADKHNHMGCPKERAETSMLEGAVLDIRGVSRIVASKEDETLY 120
DB 61 GGVKLTQSAAIRYADKHNHMGCPKERAETSMLEGAVLDIRGVSRIVASKEDETLY 120
QY 121 DFLSLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLYNDPMLDAPFLYCKF 180
DB 121 DFLSLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLYNDPMLDAPFLYCKF 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGGDHPK 218

RESULT 2

GT27_SCHMA STANDARD; PRT; 218 AA.
ID GT27_SCHMA
AC P35661;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/2)
DE antigen) (GST class-alpha).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;

RP SEQUENCE FROM N.A.
RX MEDLINE-92131046; PubMed-1775156;
RA Wright M.D., Harrison R.A., Melder A.M., Newport G.R., Mitchell G.F.;
RT "Another 26-kilodalton, glutathione S-transferase of Schistosoma
mansoni".
RL Mol. Biochem. Parasitol. 49:177-179(1991).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: M73624; -; NOT_ANNOTATED_CDS.
DR PIR: A45556; A45556.
DR HSRF: P08515; 1GTA.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C: 1.
DR Pfam: PF02798; GST_N: 1.
DR Transferrase; Antigen; Multigene family.
SQ SEQUENCE 218 AA: 25411 MW: DDD3EE9028B36185 CRC64;

Query Match 84.0%; Score 990; DB 1; Length 218;
Best Local Similarity 82.6%; Pred. No. 1.5e-81;
Matches 180; Conservative 19; Mismatches 19; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MAPKLGWKIKGLVPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTQSAAIRYADKHNHMGCPKERAETSMLEGAVLDIRGVSRIVASKEDETLY 120
DB 61 GGVKLTQSAAIRYADKHNHMGCPKERAETSMLEGAVLDIRGVSRIVASKEDETLY 120
QY 121 DFLSLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLYNDPMLDAPFLYCKF 180
DB 121 DFLSLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLYNDPMLDAPFLYCKF 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGGDHPK 218

RESULT 3

GT26_SCHMA STANDARD; PRT; 218 AA.
ID GT26_SCHMA
AC P15964;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/1)
DE antigen) (GST class-alpha).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;

RP SEQUENCE FROM N.A.
RX STRAIN-Puerto Rican;
RX MEDLINE-90348716; PubMed-2385266;
RX Troitein F., Kieny M.P., Vervaeke C., Torpier G., Pierce R.J.,
RA Ballou J.-M., Schmitt D., Lecocq J.-P., Capron A.;
RT "Molecular cloning and tissue distribution of a 26-kilodalton
Schistosoma mansoni glutathione S-transferase".
RL Mol. Biochem. Parasitol. 41:35-44(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF

```
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- TISSUE SPECIFICITY: TEGUMENT AND IN SUBTEGUMENTARY PARENCHYMAL
CC CELLS. GST 26 MAY BE ACTIVELY EXCRETED BY ADULT WORMS.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST THREE ISOENZYMES OF GST IN
CC S. MANSONI.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M31106; AAA29888.1; -.
DR EMBL: M26913; AAA29889.1; -.
DR PIR: A45523; A45523.
DR HSP: P08515; IGTA.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR Transferase; Antigen; Multigene family.
SQ SEQUENCE 218 AA; 25401 MW; 061A6548A842D6E8 CRC64;

Query Match 78.6%; Score 927; DB 1; Length 218;
Best Local Similarity 79.8%; Pred. No. 6.3e-76;
Matches 174; Conservative 15; Mismatches 29; Indels 0; Gaps 0;

QY 1 MSPILGKIKIGLVOPTRLLLEEKYEHLIERDEGDKRNKKEFLGEPNLPYIID 60
DB 1 MAPKGVKWKVGLVOPTRLLLEHELEYEERAYDNEIDAMSNDKFKLGEFPNLPYIID 60
QY 61 GDVKLTOSMAITIRYADKHNNLGGCPKRAEISMEGAVLDRIYGVSRVAYSKDEFTLV 120
DB 61 GDFKLTOSMAITIRYADKHNNLGGCPKRAEISMEGAVLDRIYGVSRVAYSKDEFTLV 120
QY 121 DFLSKPEMLKMFEDRLCKHTYLANGDHYTHDPFMDYDALDVLYMDPCLDAFPKLVFK 180
DB 121 DFLNKLPGRLKMFEDRLCKHTYLANGDHYTHDPFMDYDALDVLYMDPCLDAFPKLVFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGSDHPK 218
DB 181 KCIEDLPQIKYLNSSRYIKWPLQGWDAITFGGSDHPK 218

RESULT 4
GT29_FASHE STANDARD: PRT; 220 AA.
AC P56598;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 1 (EC 2.5.1.18) (GST1) (FHL) (GST
DE class-alpha).
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomida; Echinostomata; Fascioloidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
RT transferases of Fasciola hepatica."
RL Exp. Parasitol. 74:232-237(1992).
```

```
RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 22-220 FROM N.A.
RA Crameri S.;
RL Patent number W09008819, 09-AUG-1990.
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A00993; CAA00118.1; -.
DR HSP: P31670; IPR.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR Transferase; Antigen; Multigene family.
SQ SEQUENCE 220 AA; 25598 MW; 27B9F150B75D101F CRC64;

Query Match 60.2%; Score 709.5; DB 1; Length 220;
Best Local Similarity 60.6%; Pred. No. 1.8e-56;
Matches 129; Conservative 31; Mismatches 52; Indels 1; Gaps 1;

QY 5 LGYWKIKGLVOPTRLLLEEKYEHLIERDEGDKRNKKEFLGEPNLPYIIDGVK 64
DB 4 LGYWKIKGLVOPTRLLLEEKYEHLIERDEGDKRNKKEFLGEPNLPYIIDGVK 62
QY 65 LTOSMAITIRYADKHNNLGGCPKRAEISMEGAVLDRIYGVSRVAYSKDEFTLV 124
DB 63 LTOSMAITIRYADKHNNLGGCPKRAEISMEGAVLDRIYGVSRVAYSKDEFTLV 122
QY 125 KLPEMLKMFEDRLCKHTYLANGDHYTHDPFMDYDALDVLYMDPCLDAFPKLVCKKRIE 184
DB 123 DLPTTKMWSDFLGKPNPLRGTSVSHVDPVYEAIDAIKYLEPCLDHPNLOQFMSRIE 182
QY 185 AIPQIDKYLKSSKIYAMPLOGWQATFGGSDHPK 217
DB 183 ALPSIKAYMESNRFIKWPLNGWHAQFGGSDAPP 215

RESULT 5
GT27_FASHE STANDARD: PRT; 217 AA.
AC P31670;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GST47) (FH47)
DE (GST class-alpha).
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomida; Echinostomata; Fascioloidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;
```

RA Panaccio M., Wilson L.R., Crameri S.L., Wjffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of Fasciola hepatica.";
 RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RN ERRATUM.
 RX MEDLINE-94039664; PubMed-8224094;
 RA Panaccio M., Wilson L.R., Crameri S.L., Wjffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RN SEQUENCE OF 7-105 FROM N.A.
 RA Crameri S.;
 RL Patent number WO9008819, 09-AUG-1990.
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
 RX MEDLINE-98035725; PubMed-9367777;
 RA Rosjohn J., Fell S.C., Wilce M.C.J., Sexton J.L., Spithill T.W.,
 RA Parner M.W.;
 RT "Crystallization, structural determination and analysis of a novel
 RT parasite vaccine candidate: Fasciola hepatica glutathione
 RT S-transferase.";
 RL J. Mol. Biol. 273:857-872(1997).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M7681; AAA29140.1; -
 DR EMBL: A00996; CAA00121.1; -
 DR PDB: 1FHE; 29-JUL-98.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferase: Antigen; Multigene family; 3D-structure.
 KW INIT_MER 0 0
 FT CONFLICT 65 65 T -> I (IN REF. 3).
 FT CONFLICT 102 105 RIGF -> FEEL (IN REF. 3).
 SQ SEQUENCE 217 AA; 25281 MW; 0FB8B8FE63029E03 CRC64;
 Query Match 59.3%; Score 699; DB 1; Length 217;
 Best Local Similarity 58.2%; Pred. No. 1.5e-55;
 Matches 124; Conservative 36; Mismatches 53; Indels 0; Gaps 0;
 QY 5 LGYWKIKGLVQPTRLLEYLEEYEEHLERDEGDKMRNKKFELGEPNLPYYIDGVK 64
 DB 4 LGYWKIRGLAQPRLFLLEYLGEYEEHLGRDRKRMSEKFMGIDLPNLPYYIDDKK 63
 QY 65 LTQSMALIRYIADKHMGLGCPKPERAEISMLGCAVDIRYGSRIAYSKDFELKDYFDS 124
 DB 64 LTQSVAIMRYIADKHMGLGTPPEERARISIMIEGAANDLRIGFRVCYNPKFEVKEEYVK 123
 QY 125 KLPKMLKMFEDRLCHRTYINGDHVTHPDMLYDALDVLYMDPMCDAPFKLYCFFKRRIE 184
 DB 124 ELPKTLKMSDFLGDHRYLTGSSVSHVDEFLYETLDSIRYLAPHCLDEPFKLKEFSRIE 183
 QY 185 AIPQIDIKYKSSXYIAMPLOGMQATFGGCHPP 217
 DB 184 ALPKIKAYMESKRFTIKWPLNGMAASFQAGDAPP 216

RESULT 6
 ID GT28_FASHE STANDARD; PRT; 217 AA.
 AC P31671;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 7 (EC 2.5.1.18) (GST) (F77) (GST
 DE class-alpha).
 OS Fasciola hepatica (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fascioloidae; Fasciola.
 OX NCBI_Taxid-6192;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE-92155306; PubMed-1740183;
 RA Panaccio M., Wilson L.R., Crameri S.L., Wjffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of Fasciola hepatica.";
 RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RN ERRATUM.
 RX MEDLINE-94039664; PubMed-8224094;
 RA Panaccio M., Wilson L.R., Crameri S.L., Wjffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RN SEQUENCE OF 8-217 FROM N.A.
 RA Crameri S.;
 RL Patent number WO9008819, 09-AUG-1990.
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M7680; AAA29139.1; -
 DR EMBL: A00994; CAA00119.1; -
 DR HSP: P31670; 1FHE.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferase: Antigen; Multigene family.
 KW INIT_MER 0 0
 FT CONFLICT 35 42 NDREKWLG -> MIGNRNGNA (IN REF. 3).
 FT CONFLICT 188 196 IKYEKMSR -> SRYMSRA (IN REF. 3).
 SQ SEQUENCE 217 AA; 25196 MW; 00999ELF59E49A49E CRC64;
 Query Match 57.8%; Score 681; DB 1; Length 217;
 Best Local Similarity 58.7%; Pred. No. 6.2e-54;
 Matches 125; Conservative 29; Mismatches 59; Indels 0; Gaps 0;
 QY 5 LGYWKIKGLVQPTRLLEYLEEYEEHLERDEGDKMRNKKFELGEPNLPYYIDGVK 64
 DB 4 LGYWKIRGLAQPRLFLLEYLGEYEEHLGRDRKRMGLGFMGIDLPNLPYYIDDKK 63
 QY 65 LTQSMALIRYIADKHMGLGCPKPERAEISMLGCAVDIRYGSRIAYSKDFELKDYFDS 124
 DB 64 LTQSVAIMRYIADKHMGLGTPPEERARISIMIEGAANDLRIGFICYNPKFEELKGDYK 123

RL Acta Crystallogr. D 50:219-224(1994).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=94153886; PubMed=8110735;
 RA Ji X., Johnson W.W., Seeay M.A., Dickert L., Prasad S.M., Ammon H.L.,
 Armstrong R.N., Gilliland G.L.;
 RT "Structure and function of the xenobiotic substrate binding site of a
 glutathione S-transferase as revealed by x-ray crystallographic
 analysis of product complexes with the diastereomers of 9-(S-
 glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene.";
 RT Biochemistry 33:1043-1052(1994).
 RL
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: THE OLEFACTORY GST MAY BE CRUCIAL FOR THE ACTIVITY OF THE
 OLEFACTORY PROCESS.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MISCELLANEOUS: RAT LIVER GLUTATHIONE S-TRANSFERASES (GST) CONSIST
 OF BINARY COMBINATIONS OF THREE MAJOR CLASSES OF SUBUNITS
 DESIGNATED: YA (25 KDA), YB (27 KDA) AND YC (28 KDA).
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X04229; CAA27811.1; -
 DR EMBL: M11719; AAA41287.1; -
 DR EMBL: J02810; AAA41293.1; -
 DR PIR: A24085; A24085.
 DR PIR: A25510; A25510.
 DR PIR: A29794; A29794.
 DR PIR: S17167; S17167.
 DR PDB: 1GSB; 31-OCT-93.
 DR PDB: 1GSC; 31-OCT-93.
 DR PDB: 2GST; 31-OCT-93.
 DR PDB: 3GST; 31-JAN-94.
 DR PDB: 4GST; 31-OCT-93.
 DR PDB: 5GST; 31-OCT-93.
 DR PDB: 6GST; 08-NOV-96.
 DR PDB: 6GSU; 08-NOV-96.
 DR PDB: 6GSV; 08-NOV-96.
 DR PDB: 6GSW; 08-NOV-96.
 DR PDB: 6GSX; 08-NOV-96.
 DR PDB: 6GSY; 08-NOV-96.
 DR PDB: 5FWG; 27-JAN-99.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 KW Transferase; Multigene family; 3D-structure; Olfaction.
 FT INIT MET 0 0
 FT MUTAGEN 86 86 C->S: NO CHANGE IN ACTIVITY.
 FT CONFLICT 168 168 I -> N (IN REF. 3).
 FT CONFLICT 198 199 KS -> NC (IN REF. 2).
 FT STRAND 2 7
 FT TURN 11 13
 FT HELIX 14 22
 FT TURN 23 24
 FT STRAND 27 32
 FT TURN 37 39
 FT HELIX 43 46
 FT TURN 47 50
 FT STRAND 61 64
 FT TURN 65 66

FT STRAND 67 70
 FT HELIX 72 82
 FT TURN 83 84
 FT HELIX 90 114
 FT TURN 115 115
 FT TURN 117 118
 FT HELIX 119 128
 FT TURN 129 129
 FT HELIX 130 141
 FT TURN 142 143
 FT TURN 154 154
 FT HELIX 155 169
 FT TURN 171 176
 FT HELIX 178 188
 FT TURN 189 189
 FT HELIX 191 196
 FT TURN 197 198
 FT TURN 200 201
 FT TURN 210 211
 SQ SEQUENCE 217 AA; 25782 MW; 2ACE8D49DA785118 CRC64;
 Query Match 40.8%; Score 481.5; DB 1; Length 217;
 Best Local Similarity 44.4%; Pred. No. 4.3e-36;
 Matches 91; Conservative 38; Mismatches 71; Indels 5; Gaps 1;
 QY 4 ILGYMKIGLVPTRLLEYLEEYERDEG-----DKRNKKFELGEPNLPY 58
 DB 3 ILGYMVRGITHPIRLLELTSSYEKRAMGDADPDYRQWLNKEFKLGIDFPNLPYL 62
 QY 59 IDGDVLTOSMAIIRYIADKHNMLGGCPKRAEISMEGAVLDIRGVSRIVASKDETL 118
 DB 63 IDGSRKITQSNAMIRYLARKNHLGTEERIRADIVENQVMDNRQMLMCTNPPEKQ 122
 QY 119 KYDFLSKPEMLKMFEDRLCHTYLNGDHYTHPDFMLDALDYVLYMDPCLDAFPKLYC 178
 DB 123 KPEFLKTPKMKMLSEFLGKRPWFAGDKVTYVDFLAYDLDQYHIFPKCLDAFPKLD 182
 QY 179 FKRRIRAIPOIDKYIKSKKIAMP 203
 DB 183 FLARFGLKRTISAYMSSRYLSTPI 207
 RESULT 10
 GTM4_HUMAN STANDARD; PRT; 218 AA.
 ID GTM4_HUMAN
 AC 003013;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glutathione S-transferase Mu 4 (EC 2.5.1.18) (GSTM4-4) (GTS-Mu2)
 DE (GST class-mu 4).
 GN GSTM4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBITaxID=9606;
 RX MEDLINE=93228631; PubMed=8349586;
 RA Comstock K.E., Johnson K.J., Rifeberry D., Hennen W.D.;
 RT "Isolation and analysis of the gene and cDNA for a human Mu class
 glutathione S-transferase, GSTM4.";
 RL J. Biol. Chem. 268:16958-16965(1993).
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93228631; PubMed=8471052;
 RA Zhong S., Spurr N.K., Hayes J.D., Wolf C.R.;
 RT "Deduced amino acid sequence, gene structure and chromosomal location
 of a novel human class Mu glutathione S-transferase, GSTM4.";
 RL Biochem. J. 291:41-50(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;

QY 56 PYIDGDKLTOSMALTIRYADKHNMLGCPKERAISMLGAVLDIRYGVSHIAYSKDF 115
 DB 61 PIIIDSHKVTOSMALTIRYADKHNMLGCPKERAISMLGAVLDIRYGVSHIAYSKDF 120
 QY 116 ETLKVDLSKLPKEMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVMDPMLDAPK 175
 DB 121 EKRKPEFLGLPQDLKLYSEFLGKOPWPAODKTFPADFVLYDVLQDRHMEFPCLDAPFN 180
 QY 176 LVCFKRIEAIPOIDKYLKSKKIYAMPLOGWQATFG 211
 DB 181 LKDFMARFEGLRKISAYMKTSLRPLSPVYLKQATWG 216

RESULT 12

QY 12 GTM2_RABBIT STANDARD: PRT: 217 AA.
 AC P46409;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1995 (Rel. 38, Last annotation update)
 DE Glutathione S-transferase Mu 1 (EC 2.5.1.18) (GST Mu 1) (GST class-mu).
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95251394; PubMed=7733673;
 RA Lee S.H., Lee S.H., Han J.S., Kim Y.S., Koh J.K.;
 RT "Cloning and expression of a cDNA for mu-class glutathione S-transferase from rabbit liver."
 RL Arch. Biochem. Biophys. 318:424-429(1995).
 CC -1- FUNCTION: CONFIGURATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: WELL EXPRESSED IN RABBIT LIVER, BRAIN, AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC EMBL: L23766; AAA69665.1; -
 DR HSSP: P09488; IGTU.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PR01267; GSTRNSFRASEM.
 KW Transferase; Multigene family.
 FT INITMET 0 BY SIMILARITY.
 SQ SEQUENCE 217 AA; 25286 MW; 785EA932C543ECB2 CRC64;

Query Match 40.5%; Score 477.5; DB 1; Length 217;
 Best Local Similarity 44.6%; Pred. No. 9, 9e-36;
 Matches 95; Conservative 36; Mismatches 77; Indels 5; Gaps 1;

QY 5 LGYWKIGLVQPTRLLEYLEEKYEHLVERDEC----DKWRNKKFELGIEFPNLPYI 59
 DB 4 LGYWDVAGLALPIRMLELEYDTSYEEKKTYMGDAPNYDQSKWLSEKFTLLGDPNLPYLI 63
 QY 60 DGVKLTGSMALIRYADKHNMLGCPKERAISMLGAVLDIRYGVSHIAYSKDF 119

DB 64 DGHKLTGSMALIRYADKHNMLGCPKERAISMLGAVLDIRYGVSHIAYSKDF 123
 QY 120 VDFLSKLPKEMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVMDPMLDAPK 179
 DB 124 PEXLKLPKEMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVMDPMLDAPK 183
 QY 180 KRRIEAIPOIDKYLKSKKIYAMPLOGWQATFG 212
 DB 184 HVREGLPRISAYMKTSLRPLSPVYLKQATWG 216

RESULT 13

QY 13 GTM2_HUMAN STANDARD: PRT: 217 AA.
 AC P28161;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glutathione S-transferase Mu 2 (EC 2.5.1.18) (GSTM2-2) (GST class-mu 2).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Vorachek W.R., Pearson W.R., Rule G.S.;
 RX MEDLINE=91239584; PubMed=2034681;
 RA Ragnathan S., Chandross R.J., Kretsinger R.H., Allison T.J.,
 RT "Crystal structure of human class mu glutathione transferase GSTM2-2. Effects of lattice packing on conformational heterogeneity.";
 RL J. Mol. Biol. 238:815-832(1994).
 CC -1- FUNCTION: CONFIGURATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: THIS IS A MUSCLE ISOZYME.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC EMBL: M63509; AAA60963.1; -
 DR PIR: A39375; A39375.
 DR PDB: 1HNA; 31-JAN-94.
 DR PDB: 1HNB; 31-JAN-94.
 DR PDB: 1HNC; 31-JAN-94.
 DR PDB: 2GTU; 02-MAR-99.
 DR PDB: 3GTU; 29-JUL-99.
 DR Genew: HGNC:4634; GSTM2.
 DR MIM: 138380; -
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PR01267; GSTRNSFRASEM.
 KW Transferase; Multigene family; 3D-structure.

```

FT INIT MET 0 0
FT STRAND 2 5
FT STRAND 7 7
FT TURN 11 12
FT HELIX 13 22
FT TURN 23 24
FT STRAND 27 29
FT STRAND 32 32
FT STRAND 35 35
FT TURN 38 39
FT STRAND 41 41
FT HELIX 43 49
FT TURN 50 52
FT STRAND 61 64
FT TURN 65 66
FT STRAND 67 70
FT HELIX 72 81
FT TURN 82 84
FT HELIX 90 113
FT TURN 114 115
FT TURN 117 118
FT HELIX 119 127
FT TURN 128 129
FT HELIX 130 141
FT TURN 142 143
FT STRAND 146 146
FT TURN 147 147
FT STRAND 148 149
FT TURN 150 150
FT STRAND 151 151
FT HELIX 154 169
FT HELIX 171 176
FT HELIX 178 189
FT HELIX 191 197
FT TURN 198 198
FT TURN 214 215
SQ SEQUENCE 217 AA; 25613 MW; 25603A909482CA39 CRC64;

Query Match 40.2%; Score 473.5; DB 1; Length 217;
Best Local Similarity 44.3%; Pred. No. 2.2e-35;
Matches 94; Conservative 35; Mismatches 78; Indels 5; Gaps 1;

QY 5 LGYRKINGVLPTPLLEYLEEKEYEENHYERDEG-----DKRNRKKFELGLEFPNLPYTI 59
DB 4 LGYNNIKGLAHSHILPLLEYTDSSTEEKYRTMGADPDYDRSQGLNFKIGLDPLNPLYDI 63
QY 60 DGDVYLQSMALIRYIDRKNNMLGCGCPKERAELISMLEGAVLDIRYGSRIAYSKDFETLK 119
DB 64 DGFHKTIQSNAILRYIAKNHNLGGESEKEQIREDLLENQFMDRSQGLAKICTDPPEFKIK 123
QY 120 VDFLSKLPKMKMEDRLCHRTYTLNGDHVTHPDEMLYALDVLVLYMDPXCIDAFPKLYCF 179
DB 124 PEYLQALPEPMKIKLSQFEGKQPFELGDKITFEVDFTAYDVLERNQVFEPSCIDAFPNLKF 183
QY 180 KKRIEAIPQIDKYLKSSKYIAMPLOGMGATG 211
DB 184 ISRFEGLERKISAYMKSSRFPLRPVFTKMAVNG 215

RESULT 14
GTMU_CAVPO STANDARD; PRT; 217 AA.
AC p16413:
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glutathione S-transferase B (EC 2.5.1.18) (GST B) (GST class-mu).
GN GSTM1.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavilidae; Cavia.
NCBI_TaxID=10141;
NN [1]

```

RP SEQUENCE
 RC TISSUE=Liver:
 RX MEDLINE=90236961; Pubmed=2332413;
 RA Kamei K., Oshino R., Hara S.;
 RT "Amino acid sequence of glutathione S-transferase b from guinea pig
 liver.";
 RL J. Biochem. 107:111-117(1990).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- CATALYTIC ACTIVITY: Rx + glutathione - Hx + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY..
 DR PIR: JX0095; JX0095.
 DR HSSP: P04905; 2cst.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR03081; GST_mu.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 DR Transference: Multigene family.
 KW SEQUENCE 217 AA; 25719 MW; D29F7951DAE9365E CXC64;
 SO
 Query Match 40.2%; Score 473.5; DB 1; Length 217;
 Best Local Similarity 45.1%; Pred. No. 2.2e-35;
 Matches 92; Conservative % 34; Mismatches 73; Indels 5; Gaps 1;
 QY 5 LGYMKINGLVQPTRLLEYLEEKYEHLHYENDEG-----DKRKNKKFELGLEFPNLPYI 59
 DB 4 LGYMKIRQLTPRIILLETYNSGYEEKRYNGMDADYDRSQWLNEKFKIGDPEMLPYII 63
 QY 60 DGDVLTQSMAIIRIYADKHNMLGCGPKEAIEISMLEGAVLDIRGVSRIAVSKPFETLK 119
 DB 64 DGFHLLTQSMALIRIYARKHNLGCTEETETIRMDILENQVDIMRQGLMLCSPFEQKK 123
 QY 120 VDFLSKLPMLKMEKEDRCHKTYLNGDHYTPDPMLYALDVLVYMDPKCLDAPFKLYCF 179
 DB 124 AEFLEGIPDKKKLLSQFGKLPWFRGNKRLTYVDLAYDVLQYRLMEKCLEAFENLADF 183
 QY 180 KRIEAIPOIDIKYLSKSKYIAMPL 203
 DB 184 ISRFEGLEKISSYMKSSRFPLPKPL 207
 RESULT 15
 GTMU_CRITLO STANDARD; PRT; 217 AA.
 AC 000285;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase Y1 (EC 2.5.1.18) (Chain 3) (GST class-mu).
 OS Citriculus longicaudatus (Long-tailed hamster) (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 CC Citriculus.
 CC NCBI_Taxid:10030;
 RX MEDLINE=91031445; Pubmed=2226437;
 RA de Saint Vincent B.R., Hyrien O., Debatisse M., Buttin G.;
 RT "Complication of mu class glutathione S-transferase genes and an
 adenylate deaminase gene in colormycin-resistant Chinese hamster
 fibroblasts.";
 RL Eur. J. Biochem. 193:19-24(1990).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- CATALYTIC ACTIVITY: Rx + glutathione - Hx + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 10:12:24 ; Search time 15.0413 Seconds

(without alignments)
2986.328 Million cell updates/sec

Title: US-10-081-408-4

Sequence: 1 MSPILGWIKIGLVQPTRL.....IAMPLQGMQATFGGDHPPK 218

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1049	89.0	218	5	Q94745 schistosoma
2	718	60.9	218	5	Q25595 clonorchis
3	698	59.2	218	5	Q9XYL9 09xyl9 fasciola gl
4	512.5	43.5	218	6	Q9N0V4 09n0v4 bos taurus
5	505.5	42.9	219	13	Q90WM9 090wm9 xenopus lae
6	502.5	42.6	223	5	Q97117 097117 boophilus m
7	500.5	42.5	218	6	Q9R516 09r516 mus musculu
8	477.5	40.3	218	6	Q9TSM5 09tasm5 macaca fasc
9	475	40.3	221	6	Q9EBB0 09ebb0 macaca fusc
10	474.5	40.2	218	6	Q9TSM4 09tasm4 macaca fasc
11	474.5	40.2	219	5	Q27653 027653 echinococu
12	473.5	40.2	219	5	Q16058 016058 echinococu
13	471.5	40.0	218	11	Q91Y83 091y83 cavia porce
14	464.5	39.4	218	11	Q9WU21 09wu21 rattus norv
15	456.5	38.7	218	11	Q9DD25 09dd25 mus musculu
16	456.5	38.7	218	11	Q9DCB8 09dcb8 mus musculu

17	448.5	38.0	225	4	Q96HA3 096ha3 homo sapien
18	442.5	37.5	225	6	Q9BEA9 09bea9 macaca fusc
19	436.5	37.0	195	4	Q05465 005465 homo sapien
20	434.5	36.9	225	11	Q921B2 0921b2 rattus norv
21	434.5	36.9	232	11	Q9D5T6 09d5t6 mus musculu
22	433.5	36.8	220	4	Q8WWE1 08wwe1 homo sapien
23	407.5	34.6	188	6	Q9M2B4 09m2b4 capra hircu
24	377.5	32.0	219	5	Q9U582 09u582 psoroptes o
25	347	29.4	181	4	Q8TC98 08tc98 homo sapien
26	319.5	27.1	220	5	Q8RTB4 08rtb4 boophilus m
27	288.5	24.5	125	6	Q2S583 02s583 sus scrofa
28	256.5	21.8	208	5	Q02636 002636 brugia mala
29	254.5	21.6	208	13	Q9DDU5 09ddu5 brachydantio
30	247.5	21.0	208	5	Q27711 027711 onchocerca
31	240	20.4	210	4	Q00460 000460 homo sapien
32	240	20.4	210	4	Q15690 015690 homo sapien
33	236	20.0	180	5	Q9NHB2 09nhb2 mytilus edu
34	233	20.0	209	6	Q9TTY8 09tty8 capra hircu
35	213.5	18.1	216	13	Q9W647 09w647 oncorhynch
36	212.5	18.0	82	4	Q9UE37 09ue37 homo sapien
37	193.5	16.4	221	11	Q9DCU1 09dcu1 mus musculu
38	192.5	16.3	208	5	P91505 091505 caenorhabd
39	191.5	16.2	226	11	Q9TIX3 09tix3 rattus norv
40	189.5	16.1	221	11	P70686 070686 mesocricetu
41	187	15.9	210	5	Q9N4X8 09n4x8 caenorhabd
42	186.5	15.8	222	6	Q9N4J6 09n4j6 ovis aries
43	185.5	15.7	198	5	Q8TJ39 08tj39 dictyostell
44	183.5	15.6	210	5	Q966G8 0966g8 caenorhabd
45	182.5	15.5	208	5	Q61750 061750 caenorhabd

ALIGNMENTS

RESULT 1	Q94745	PRELIMINARY;	PRT;	218 AA.
ID	Q94745			
AC	Q94745;			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	26kd glutathione S-transferase.			
OS	Schistosoma mekongi.			
OC	Schistosomatidae; Platyhelminthes; Trematoda; Digenea; Strigoidae;			
OC	Eukaryota; Metazoa; Schistosomatidae; Schistosoma.			
OX	NCBI_TaxID=38744;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=THAILAND;			
RX	MEDLINE=98222395; PubMed=9561610;			
RA	Grams S.V., Grams R., Korge G., Vivanant V., Upatham S.;			
RT	"Cloning and sequencing of the 26 kDa glutathione S-transferase gene			
RT	of Schistosoma mekongi".			
RL	Southeast Asian J. Trop. Med. Public Health 28:570-574(1997).			
DR	EMBL: 107663; CAA68944.1; .			
DR	HSSP: P08515; IGTa.			
DR	InterPro: IPR004046; GST_Cterm.			
DR	InterPro: IPR004045; GST_Nterm.			
DR	Pfam: PF00043; GST_C: 1.			
DR	Pfam: PF02798; GST_N: 1.			
KW	Transferase.			
SQ	SEQUENCE 218 AA; 25486 MW; D3080620B19DE23A CRC64;			
Query Match	89.0%; Score 1049; DB 5; Length 218;			
Best Local Similarity	89.4%; Pred. No. 1.5e-86;			
Matches 195; Conservative 11; Mismatches 12; Indels 0; Gaps 0;				
OY	1 MSPILGWIKIGLVQPTRLLEYLEEYERDEGDKWRNKKFELGLEFPNLPYYID 60			
I				
DB	1 MAPILGWIKIGLVQPTRLLEYLEEYERDEGDKWRNKKFELGLEFPNLPYYID 60			
I				
OY	61 GQVRLQSMALIRYIADKHKMLGCGPKERADISLBEAVIADIRGVSRIRYNSDFELTKY 120			
I				

Db 61 GVKLTGSMALIRYIADKHNHMGSGKERAEITMLEGAVSDIRSGVSIAYNKDEFETLKV 120
 QY 121 DFLSKLEPMLEKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVCFK 180
 121 DFLKLEPMLEKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLLKSSKYIAMPLOGMOATFGGDDHPK 218
 181 KRIENIPQINEYLRSSKYIEMPLQGMQATFGGDDHPK 218

RESULT 2

Q25595 ID Q25595 PRELIMINARY; PRT; 218 AA.
 AC 025595;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Putative glutathione transferase.
 OS Clonorchis sinensis.
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Opisthorchiida; Opisthorchiata; Opisthorchioidea; Opisthorchiidae;
 OC Clonorchis.
 NC NCB1_TaxID=79923;
 RX NCBI_TaxID=79923;
 RP SEQUENCE FROM N.A.
 RA Hong S.-J., Lee D.-H.;
 RT "Cloning and over expression of 26 kDa glutathione S-transferase from
 Clonorchis sinensis."
 RL Thesis (1996), Parasitology, Chung-Ang University College of Medicine,
 Seoul.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hong S.-J.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: L47992; AAB46369.3; -.
 DR HSSP: P31670; 1FHE.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferase.
 KW TRANSFERASE.
 SQ SEQUENCE 218 AA; 25038 MW; 7CB17C7B837A0B7C CRC64;

Query Match 60.9%; Score 718; DB 5; Length 218;
 Best Local Similarity 59.2%; Pred. No. 9.1e-57;
 Matches 125; Conservative 37; Mismatches 52; Indels 0; Gaps 0;

QY 1 MSPILGYWKIRGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
 1 MAPVLGYWKIRGLAOPRILLLEYVGDSEHSGRCDEKQNDKHNKLGLELPNLPYYKD 60
 Db 1
 QY 61 GVKLTGSMALIRYIADKHNHMGSGKERAEITMLEGAVSDIRSGVSIAYNKDEFETLKV 120
 121 DFLSKLEPMLEKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVCFK 180
 121 DFLKLEPMLEKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVCFK 180
 Db 61 GVKLTGSMALIRYIADKHNHMGSGKERAEITMLEGAVSDIRSGVSIAYNKDEFETLKV 120
 121 DFLSKLEPMLEKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVCFK 180
 121 DFLKLEPMLEKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLLKSSKYIAMPLOGMOATFGGDDHPK 218
 181 KRIENIPQINEYLRSSKYIEMPLQGMQATFGGDDHPK 218
 Db 181 KRIEAIPOIDKYLLKSSKYIAMPLOGMOATFGGDDHPK 218
 181 KRIENIPQINEYLRSSKYIEMPLQGMQATFGGDDHPK 218

RESULT 3

Q9XYL9 ID Q9XYL9 PRELIMINARY; PRT; 218 AA.
 AC 09XYL9;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase.

GN GST-1.
 OS Fasciola gigantica.
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fascioloidae; Fasciolidae; Fasciola.
 NC NCB1_TaxID=46835;
 RX NCBI_TaxID=46835;
 RP SEQUENCE FROM N.A.
 RA STRAIN=THAILAND.
 RA Grams S.V., Grams R., Sobhon P., Vaynant V., Upatham E.S.;
 RT "Molecular cloning of expressed antigens from Fasciola gigantica."
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF112567; AAD23997.1; -.
 DR HSSP: P31670; 1FHE.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferase.
 KW TRANSFERASE.
 SQ SEQUENCE 218 AA; 25337 MW; E9423D75C3F2EEAF CRC64;

Query Match 59.2%; Score 698; DB 5; Length 218;
 Best Local Similarity 58.7%; Pred. No. 5.8e-55;
 Matches 125; Conservative 35; Mismatches 53; Indels 0; Gaps 0;

QY 5 LGYWKIRGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDGVK 64
 5 LGYWKIRGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDGVK 64
 Db 5
 QY 65 LTQSMALIRYIADKHNHMGSGKERAEITMLEGAVSDIRSGVSIAYNKDEFETLKV 124
 65 LTQSMALIRYIADKHNHMGSGKERAEITMLEGAVSDIRSGVSIAYNKDEFETLKV 124
 Db 65
 QY 125 KLPEMLKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVCFK 184
 125 ELPEMLKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVCFK 184
 Db 125
 QY 185 AIPOIDKYLLKSSKYIAMPLOGMOATFGGDDHPK 217
 185 AIPOIDKYLLKSSKYIAMPLOGMOATFGGDDHPK 217
 Db 185
 QY 185 DLPRIKAYMESEKFIKMPNLNMTWTSFGGDDAP 217
 185 DLPRIKAYMESEKFIKMPNLNMTWTSFGGDDAP 217

RESULT 4

Q9NOV4 ID Q9NOV4 PRELIMINARY; PRT; 218 AA.
 AC 09NOV4;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Class mu glutathione S-transferase.
 GN GSTM.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NC NCB1_TaxID=9913;
 RX NCBI_TaxID=9913;
 RP SEQUENCE FROM N.A.
 RA Tissue=LENS.
 RA Jimenez-Arenas J.V., Garland D.;
 RT "A lens glutathione S-transferase, class mu, with thiol-specific
 antioxidant activity."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF249588; AAF64308.1; -.
 DR HSSP: P09488; 1GTU.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PR01267; GSTRNSFRASEM.
 DR Transferase.
 KW TRANSFERASE.
 SQ SEQUENCE 218 AA; 25635 MW; 3D02EA0F43C07B0A CRC64;

Best Local Similarity 44.9%, Pred. No. 3,4e-37;
Matches 97; Conservative 39; Mismatches 75; Indels 5; Gaps 1;

```

QY 1 MSPILGYWKIRKGLVQPTLLLEYLEEKYEHLERDEG-----DKWRNKKFELGLEFPNL 55
DB 1 MPMILGYWDIRGLAIAIRLLLEYTGSSYEKKRYTGWADPDYDSQWLNEKFKLGIDFPNL 60
QY 56 PYITGDVKTLOSMAIIRYADKNHMLGCGPKERAELISLGEAVLDIRGYSRIRAYSKDF 115
DB 61 PYLIDGTHKITQSNAILRYARKHNLGCEETEKEKIRVDILENQADVSNQRLARVCYSPDF 120
QY 116 ETLKVDPLSKLPEMLKMFEDRLCHKTLYLNGDHTHPDFMLYDALDVLYMDPMLCDAPFK 175
DB 121 EKLKVEYLEQJLPGWVKLSQFLGQRTWVGEKITYVDPLAYIIDLHLIFEFTCLDAPFN 180
QY 176 LVCFKKRIEALPQIDKYKSSKYIAMPLOGQWATFG 211
DB 181 LKDFVAREVFLKRIASVYKTSRFLRTPLYTKVATWG 216

RESULT 8
Q9TSM5 PRELIMINARY; PRT: 218 AA.
AC Q9TSM5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glutathione S-transferase mu-class subunit M1 (EC 2.5.1.18).
GN GSTM1.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=LIVER;
RC MEDLINE=20330602; PubMed=10869451;
RA Wang C., Bammaler T.K., Guo Y., Kelly E.J., Eaton D.L.;
RT "Mu-class GSTs are responsible for aflatoxin B(1)-8, 9-epoxide-
RT conjugating activity in the nonhuman primate macaca fascicularis
RT liver.";
RL Toxicol. Sci. 56:26-36(2000).
DR EMBL; AF200709; AAF08539.1; -.
DR HSSP; P09488; 1GUU.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR003081; GST_mu.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR PRINTS; PRO1267; GSTRNSFRASEM.
KW Transferase
SEQUENCE 218 AA; 25577 MW; EESBAD80F6C95EB CRC64;

```

Query Match 40.5%; Score 477.5; DB 6; Length 218;
Best Local Similarity 43.5%; Pred. No. 3,9e-35;
Matches 94; Conservative 36; Mismatches 81; Indels 5; Gaps 1;

```

QY 1 MSPILGYWKIRKGLVQPTLLLEYLEEKYEHLERDEG-----DKWRNKKFELGLEFPNL 55
DB 1 MPMILGYWDIRGLAIAIRLLLEYTGSSYEKKRYTGWADPDYDSQWLNEKFKLGIDFPNL 60
QY 56 PYITGDVKTLOSMAIIRYADKNHMLGCGPKERAELISLGEAVLDIRGYSRIRAYSKDF 115
DB 61 PYLIDGTHKITQSNAILRYARKHNLGCEETEKEKIRVDILENQADVSNQRLARVCYSPDF 120
QY 116 ETLKVDPLSKLPEMLKMFEDRLCHKTLYLNGDHTHPDFMLYDALDVLYMDPMLCDAPFK 175
DB 121 EKLKVEYLEQJLPGWVKLSQFLGQRTWVGEKITYVDPLAYIIDLHLIFEFTCLDAPFN 180
QY 176 LVCFKKRIEALPQIDKYKSSKYIAMPLOGQWATFG 211
DB 181 LKDFISHFEGLEKISAVYKSSRFLPKPLYTRVAVWG 216

```

RESULT 9
Q9EBR0 PRELIMINARY; PRT: 221 AA.

```

AC Q9EBR0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glutathione transferase M2.
OS Macaca fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9542;
RN [1]
RP SEQUENCE FROM N.A.
RA Beuckmann C.T., Fujimori K., Urade Y.;
RT "Macaca fuscata glutathione transferase M2.";
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB025799; BAB40442.1; -.
DR HSSP; P28161; 1HNA.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR003081; GST_mu.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR PRINTS; PRO1267; GSTRNSFRASEM.
KW Transferase
SEQUENCE 221 AA; 26068 MW; 01F22BACDA57FE9 CRC64;

```

Query Match 40.3%; Score 475; DB 6; Length 221;
Best Local Similarity 44.7%; Pred. No. 6,7e-35;
Matches 97; Conservative 35; Mismatches 79; Indels 6; Gaps 2;

```

QY 1 MSPILGYWKIRKGLVQPTLLLEYLEEKYEHLERDEG-----DKWRNKKFELGLEFPNL 54
DB 3 MPMILGYWDIRGLAIAIRLLLEYTGSSYEKKRYTGWADPDYDSQWLNEKFKLGIDFPNL 62
QY 55 LPYITGDVKTLOSMAIIRYADKNHMLGCGPKERAELISLGEAVLDIRGYSRIRAYSKDF 114
DB 63 PYLIDGTHKITQSNAILRYARKHNLGCEETEKEKIRVDILENQADVSNQRLARVCYSPDF 122
QY 115 ETLKVDPLSKLPEMLKMFEDRLCHKTLYLNGDHTHPDFMLYDALDVLYMDPMLCDAPFK 174
DB 123 EKLKVEYLEQJLPGWVKLSQFLGQRTWVGEKITYVDPLAYIIDLHLIFEFTCLDAPFN 182
QY 175 LVCFKKRIEALPQIDKYKSSKYIAMPLOGQWATFG 211
DB 183 LKDFISHFEGLEKISAVYKSSRFLPKPLYTRVAVWG 219

```

RESULT 10
Q9TSM4 PRELIMINARY; PRT: 218 AA.

```

AC Q9TSM4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glutathione S-transferase mu-class subunit M2 (EC 2.5.1.18).
GN GSTM2.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=LIVER;
RC MEDLINE=20330602; PubMed=10869451;
RA Wang C., Bammaler T.K., Guo Y., Kelly E.J., Eaton D.L.;
RT "Mu-class GSTs are responsible for aflatoxin B(1)-8, 9-epoxide-
RT conjugating activity in the nonhuman primate macaca fascicularis

```

RT liver ";
 RL Toxicol. Sci. 56:26-36(2000).
 DR EMBL: AF2007710; AAF08540.1; -.
 DR HSSP: P28161; 1HNA.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTNRSFRASEM.
 DR Transferase.
 KW SEQUENCE 218 AA; 25708 MW; F2E509C3949F9051 CRC64;

Query Match 40.2%; Score 474.5; DB 5; Length 218;
 Best Local Similarity 42.4%; Pred. No. 7,3e-35;
 Matches 96; Conservative 34; Mismatches 81; Indels 5; Gaps 1;
 OY 1 MSPILGWKIKGLVOPTRLLLEYLEKEYEHLERDEG-----DKWRNKKFELGLEFPNL 55
 DB 1 MPMTGLGWNIRGLASIRLLLEYTSSYEKKYTMGDADYDRSOWLNKFKLGLDFPNL 60
 OY 56 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKITQSNALIRYIARKHNLCGETEKEKIREDLLENQMLDMRQMLARLCYDPDF 120
 OY 116 ETLKVDLSKLPENLMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVYMDPCLDAFPK 175
 DB 121 EKLKPEYLLGLPEMLKLYSQFLGDKITFDYDFVGLERNNOVFEPSCLDAFPN 180
 OY 176 LVCFKKRIEAIPOIDKYLKSSKIYAMPLOGMOATFG 211
 DB 181 LKDFISREGLERISAYMKSSRLPRPVFTKMAVWG 216

RESULT 11
 ID 027653 PRELIMINARY; PRT; 219 AA.
 AC 027653;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glutathione transferase (EC 2.5.1.18).
 GN EMGST1.
 OS Echinococcus multilocularis.
 OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
 OC Cyclophyllidae; Taeniidae; Echinococcus.
 OX NCBI_TaxID=6211;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96379220; PubMed=8784771;
 RA Liebau E., Muller V., Lucius R., Walter R.D., Henkle-Duhrsen K.;
 RT "Molecular cloning, expression and characterization of a recombinant
 glutathione S-transferase from Echinococcus multilocularis.";
 RL Mol. Biochem. Parasitol. 77:49-56(1996).
 DR EMBL: X85736; CAA59739.1; -
 DR HSSP: P20136; 1GST.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferase.
 KW SEQUENCE 219 AA; 25470 MW; 74FA820AC56F745B CRC64;

Query Match 40.2%; Score 474.5; DB 5; Length 219;
 Best Local Similarity 42.4%; Pred. No. 7,4e-35;
 Matches 92; Conservative 40; Mismatches 80; Indels 5; Gaps 1;
 OY 1 MSPILGWKIKGLVOPTRLLLEYLEKEYEHLERDEGDKWRNKKFELGLEFPNL 55
 DB 1 MPMTGLGWNIRGLASIRLLLEYTSSYEKKYTMGDADYDRSOWLNKFKLGLDFPNL 60
 OY 56 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKITQSNALIRYIARKHNLCGETEKEKIREDLLENQMLDMRQMLARLCYDPDF 120

DB 61 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERRARVLMHLOCEVVDLRMAFTRTCYSPDF 120
 OY 116 ETLKVDLSKLPENLMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVYMDPCLDAFPK 175
 DB 121 EKLKPEYLLGLPEMLKLYSQFLGDKITFDYDFVGLERNNOVFEPSCLDAFPN 180
 OY 176 LVCFKKRIEAIPOIDKYLKSSKIYAMPLOGMOATFG 212
 DB 181 LKAYLSREFNLPALDYMASKKEFKTCPCNGASAKMWG 217

RESULT 12
 ID 016058 PRELIMINARY; PRT; 219 AA.
 AC 016058;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase.
 GN GSTMU.
 OS Echinococcus granulosus.
 OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
 OC Cyclophyllidae; Taeniidae; Echinococcus.
 OX NCBI_TaxID=6210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fernandez V., Zaha A., Fernandez C.;
 RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Fernandez Mancebo V., Chalar C., Martinez C., Zaha A.,
 RA Fernandez Granja C.;
 RT "EGST: a glutathione S-transferase gene from Echinococcus
 granulosus.";
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF005928; AAB6318.1; -
 DR EMBL: AF101269; AAD16438.1; -
 DR HSSP: P20136; 1GST.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferase.
 KW SEQUENCE 219 AA; 25553 MW; 8883E70AD075D154 CRC64;

Query Match 40.2%; Score 473.5; DB 5; Length 219;
 Best Local Similarity 42.4%; Pred. No. 9e-35;
 Matches 92; Conservative 40; Mismatches 80; Indels 5; Gaps 1;

OY 1 MSPILGWKIKGLVOPTRLLLEYLEKEYEHLERDEGDKWRNKKFELGLEFPNL 55
 DB 1 MPMTGLGWNIRGLASIRLLLEYTSSYEKKYTMGDADYDRSOWLNKFKLGLDFPNL 60
 OY 56 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERRARVLMHLOCEVVDLRMAFTRTCYSPDF 120
 OY 116 ETLKVDLSKLPENLMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVYMDPCLDAFPK 175
 DB 121 EKLKPEYLLGLPEMLKLYSQFLGDKITFDYDFVGLERNNOVFEPSCLDAFPN 180
 OY 176 LVCFKKRIEAIPOIDKYLKSSKIYAMPLOGMOATFG 212
 DB 181 LKAYLSREFNLPALDYMASKKEFKTRPCNGASAKMWG 217

RESULT 13
 ID 091Y83 PRELIMINARY; PRT; 218 AA.
 AC 091Y83;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Glutathione S-transferase subunit gyc.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
 OX NCBI_TaxID=10141;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-HARTLEY; TISSUE=LIVER;
 RX MEDLINE=98297271; PubMed=9633615;
 RA Hiratsuka A., Ogura K., Fujioka H., Sakamoto Y., Okuda H., Wada K.,
 Tanaka T., Nishiyama T., Matabe T.;
 RT "Guinea pig liver Mu-class glutathione S-transferase M1-2 cross-reacts
 with antibodies to both rat Mu- and theta-class glutathione S-
 transferases.";
 RT
 RL Arch. Biochem. Biophys. 354:188-196(1998).
 DR EMBL: AB000448; BAB47185.1;
 DR InterPro: IPR004046; GST_Cterm.
 DR Pfam: PF00043; GST_C.1.
 DR Pfam: PF02798; GST_N.1.
 DR Transferase.
 KM
 SQ SEQUENCE 218 AA; 25695 MW; D2B9E31F9FEF1B8 CRC64;
 Query Match 40.0%; Score 471.5; DB 11; Length 218;
 Best Local Similarity 44.7%; Pred. No. 1.4e-34;
 Matches 93; Conservative 33; Mismatches 77; Indels 5; Gaps 1;
 QY 1 MSPIIGYMKIKGLVOTRLLLEYLEKYEHLERDEG-----DKRNKKFELGLEPPL 55
 DB 1 MPTLTGYMDIRGLTNAIRLLLEDTSYEKRYTMGDADFTSQWLNRKFKGLDFPML 60
 QY 56 PYIDGDVVLQTOSMAIRYIADKHNMLGCGPKERAISMLEGAVLDIRGVSRIAYSKDF 115
 DB 61 PVLIDGTHLQTSNMLRIARLNKNCGTEEBRIMDLNOMDFRQLQGNVCSPPF 120
 QY 116 ETLKVDLSKLPDLKMFEDRLCHKTYLNGDHYHDFMLYDALDVLVYMDPMDLAPFK 175
 DB 121 EKKKEPEFLKSDPKMKKLSQFLGEPWFAGNKRITFVDLAYVDLQHRIFEPKCLEAFPN 180
 QY 176 IVCFKRIRAIPOIDKYLKSSKIYAMPL 203
 DB 181 LKEFTSRFEGKLKISAYMKSSSFLPPPM 208
 RESULT 14
 Q9WU21 PRELIMINARY; PRT: 218 AA.
 AC Q9WU21.
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase Yb4 (EC 2.5.1.18).
 GN GSTRYB4.
 OS Rattus norvegicus (rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY;
 RX MEDLINE=88298790; PubMed=3403534;
 RA Lai H.C., Qian B., Grove G., Tu C.P.;
 RT "Gene expression in the evolution of the yb multigene family.";
 RT "Gene expression in the evolution of the yb multigene family.";
 RL J. Biol. Chem. 263:11389-11395(1988).
 DR EMBL: AF106661; AAD22630.1;
 DR HSSP: P04905; ZGST.

DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C.1.
 DR Pfam: PF02798; GST_N.1.
 DR Transferase.
 KM
 SQ SEQUENCE 218 AA; 25643 MW; 4E9C22P9AEAC6C4 CRC64;
 Query Match 39.4%; Score 464.5; DB 11; Length 218;
 Best Local Similarity 43.3%; Pred. No. 5.8e-34;
 Matches 90; Conservative 35; Mismatches 78; Indels 5; Gaps 1;
 QY 1 MSPIIGYMKIKGLVOTRLLLEYLEKYEHLERDEG-----DKRNKKFELGLEPPL 55
 DB 1 MAMITLGMVNRGLTHPIRLLLEDTSDSYEKRYTMGDADNFRSQWLSEKFNGLDIPML 60
 QY 56 PYIDGDVVLQTOSMAIRYIADKHNMLGCGPKERAISMLEGAVLDIRGVSRIAYSKDF 115
 DB 61 PVLIDGSHKVTOSNMLRIARLNKNCGTEEBRIVDTLNOVMDTRHLHIVCCSPDF 120
 QY 116 ETLKVDLSKLPDLKMFEDRLCHKTYLNGDHYHDFMLYDALDVLVYMDPMDLAPFK 175
 DB 121 EKOKPEFLKSIPEKMKIYSEFLGKRPWFAGDKVTYVDLAYVDLDDYRMFEPECLDAPFN 180
 QY 176 IVCFKRIRAIPOIDKYLKSSKIYAMPL 203
 DB 181 LKDFLARFEGKLKISAYMKSSSFLPPPV 208
 RESULT 15
 Q9DD25 PRELIMINARY; PRT: 218 AA.
 AC Q9DD25.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 0610005A07Rik protein.
 GN RPRGRIPI OR 0610005A07Rik.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schiomi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo N.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Winking L.,
 Wyshnaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690(2001).
 RL EMBL: AK002213; BAB21939.1;
 DR EMBL: P28161; ZGTU
 DR HSSP: MGI:1915562; 0610005A07Rik.
 DR MGI:1915562; 0610005A07Rik.
 DR MGI:1932134; Rprgripl.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C.1.

DR N-PSDB; AAV99373.
 XX
 PT New bone morphogenetic fusion proteins - comprising a purification
 PT tag and a bone morphogenetic active fragment, used for enhancing
 PT wound healing or bone growth
 PS
 XX Disclosure; Page 39-40; 64pp; English.
 CC
 CC The present sequence represents a protein comprising a purification tag
 CC that was used in the creation of the bone morphogenetic fusion proteins
 CC of the invention. The bone morphogenetic fusion protein may contain some
 CC or all of the following elements: a purification tag, a proteinase site,
 CC an ECM/bone binding site, a second proteinase site, and a bone
 CC morphogenetic protein active fragment. The fusion proteins of the
 CC invention also includes proteins that have transforming growth factor
 CC beta active fragments instead of bone morphogenetic fusion active
 CC fragments. The bone morphogenetic fusion proteins can be used for
 CC enhancing wound healing or bone growth.
 CC
 SQ Sequence 218 AA;
 Query Match 98.7%; Score 1149; DB 20; Length 218;
 Best Local Similarity 98.6%; Pred. No. 6.6e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVPTRLLEYLEEKEEHLXERDEGDKMNKKFELGLEPNNLPYYID 60
 DB 1 MSPILGYWKIKGLVPTRLLEYLEEKEEHLXERDEGDKMNKKFELGLEPNNLPYYID 60
 QY 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRIVASKDETLKV 120
 DB 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRIVASKDETLKV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCLDAFPRLVCFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLVYMDPCLDAFPRLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGMQATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGMQATFGGSDHPK 218
 RESULT 2
 ABB09809
 ID ABB09809 standard; Protein: 218 AA.
 AC ABB09809;
 XX
 AC 22-JUL-2002 (first entry)
 DT
 XX Glutathione-S-transferase (GST) of Schistosoma japonicum.
 DE
 XX Glutathione-S-transferase; GST; enzyme; metallothionein; MT-1IA;
 KW heavy metal contact; UV light.
 KM
 XX Schistosoma japonicum.
 OS
 XX FR2813529-A1.
 PN
 XX 08-MAR-2002.
 PD
 XX 31-JUL-2001; 2001FR-0010239.
 PF
 XX 06-SEP-2000; 2000ES-0000217.
 PR
 XX (PROV-) PROVITAL SA.
 PA
 XX Armengol SR, Benalagues MA;
 PI
 XX MPI, 2002-332437/37.
 DR
 XX Topical compositions, used to protect against heavy metals and
 PT ultraviolet radiation, contain metallothionein proteins -

XX
 PS Example; Page 28-29; 33pp; French.
 XX
 CC The present sequence represents the glutathione-S-transferase (GST)
 CC protein of Schistosoma japonicum. A human metallothionein class IIA
 CC (MT-1IA) was expressed fused to the present GST protein, for
 CC purification purposes. The MT-1IA protein was used to produce
 CC topical pharmaceutical and cosmetic compositions, in the form of gels,
 CC creams, ointments, soaps and body lotions. In the compositions, the
 CC metallothionein proteins have a prosthetic group attached to a zinc
 CC ion. The topical compositions can be used for protection of human or
 CC animal external tissue against the effects of heavy metal contact and
 CC UV light.
 CC
 SQ Sequence 218 AA;
 Query Match 98.7%; Score 1149; DB 23; Length 218;
 Best Local Similarity 98.6%; Pred. No. 6.6e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVPTRLLEYLEEKEEHLXERDEGDKMNKKFELGLEPNNLPYYID 60
 DB 1 MSPILGYWKIKGLVPTRLLEYLEEKEEHLXERDEGDKMNKKFELGLEPNNLPYYID 60
 QY 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRIVASKDETLKV 120
 DB 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRIVASKDETLKV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCLDAFPRLVCFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLVYMDPCLDAFPRLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGMQATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGMQATFGGSDHPK 218
 RESULT 3
 AAM02122
 ID AAM02122 standard; Protein: 224 AA.
 AC AAM02122;
 XX
 AC 28-OCT-1996 (first entry)
 DT
 XX Glutathione-S-transferase epitope.
 DE
 XX Bone resorption disease; osteoporosis; src SH2 domain antagonist;
 KW src homology 2 domain; glutathione-S-transferase; fyn SH2;
 KM Grb2 SH2; SH-PTP2 SH2.
 KW
 XX Not specified.
 OS
 XX EP727211-A1.
 PN
 XX 21-AUG-1996.
 PD
 XX 07-FEB-1996; 96EP-0200270.
 PF
 XX 29-DEC-1995; 95US-0580868.
 PR
 XX 10-FEB-1995; 95US-0386381.
 PR
 XX 07-MAR-1995; 95US-0400220.
 PR
 XX 30-JUN-1995; 95US-0497357.
 PR
 XX 11-OCT-1995; 95US-0541080.
 PA
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PI
 XX Dunnington DJ;
 XX MPI, 1996-372674/38.
 DR
 XX Use of selective src SH2 domain ligand - to prepare medicament for
 PT treating bone resorption disease

```
XX Example 11; Page 31-32; 47pp; English.
PS
XX A glutathione-s-transferase epitope (AAW02122) was used to construct
CC fusion proteins with human SH2 domains from fyn (see also AAW02124),
CC Grb2 (AAW02127) and SH-PTP2 (AAW02125); a similar epitope (AAW02123) was
CC used to construct a fusion protein with human p85 SH2 domain (see
CC also AAW02126). These fusion proteins, and tagged proteins including
CC other SH2 domains (see also AAW02119-21), can be used in binding
CC assays to determine the specificity of cpds. to inhibit SH2 domains;
CC cpds. that selectively inhibit the human src SH2 domain are useful
CC in treating bone resorption diseases such as osteoporosis.
XX
SQ Sequence 224 AA;
Query Match 98.7%; Score 1149; DB 17; Length 224;
Best Local Similarity 98.6%; Pred. No. 6.9e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEENHYERDEGDKMKNKKFELGLEFPNLPYYTD 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEENHYERDEGDKMKNKKFELGLEFPNLPYYTD 60
OY 61 GDVKLTQSMATIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRVAYSKEDETLKV 120
DB 61 GDVKLTQSMATIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRVAYSKEDETLKV 120
OY 121 DFLSKLPKMKFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCKLDAFPKLVCFK 180
DB 121 DFLSKLPKMKFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCKLDAFPKLVCFK 180
OY 121 DFLSKLPKMKFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCKLDAFPKLVCFK 180
DB 121 DFLSKLPKMKFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCKLDAFPKLVCFK 180
OY 181 KRIEAIPQIDKYLKSSKTYAMPLOGWQATFGGDDHPK 218
DB 181 KRIEAIPQIDKYLKSSKTYAMPLOGWQATFGGDDHPK 218

RESULT 4
AAW02123
ID AAW02123 standard; Protein: 225 AA.
XX
AC AAW02123;
XX
DT 28-OCT-1996 (first entry)
XX
DE Glutathione-s-transferase epitope.
XX
KW Bone resorption disease; osteoporosis; src SH2 domain antagonist;
KW src homology 2 domain; glutathione-s-transferase; p85 SH2.
XX
OS Not specified.
XX
PN EP727211-A1.
XX
PD 21-AUG-1996.
XX
PF 07-FEB-1996; 96EP-0200270.
XX
PR 29-DEC-1995; 95US-0580868.
PR 10-FEB-1995; 95US-0386381.
PR 07-MAR-1995; 95US-0400220.
PR 30-JUN-1995; 95US-0497357.
PR 11-OCT-1995; 95US-0541080.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI Dunnington DJ;
XX
WPI: 1996-372674/38.
XX
DR Use of selective src SH2 domain ligand - to prepare medicament for
PT treating bone resorption disease
XX
PS Example 11; Page 31-32; 47pp; English.
```

```
XX A glutathione-s-transferase epitope (AAW02122) was used to construct
CC fusion proteins with human SH2 domains from fyn (see also AAW02124),
CC Grb2 (AAW02127) and SH-PTP2 (AAW02125); a similar epitope (AAW02123) was
CC used to construct a fusion protein with human p85 SH2 domain (see
CC also AAW02126). These fusion proteins, and tagged proteins including
CC other SH2 domains (see also AAW02119-21), can be used in binding
CC assays to determine the specificity of cpds. to inhibit SH2 domains;
CC cpds. that selectively inhibit the human src SH2 domain are useful
CC in treating bone resorption diseases such as osteoporosis.
XX
SQ Sequence 225 AA;
Query Match 98.7%; Score 1149; DB 17; Length 225;
Best Local Similarity 98.6%; Pred. No. 6.9e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEENHYERDEGDKMKNKKFELGLEFPNLPYYTD 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEENHYERDEGDKMKNKKFELGLEFPNLPYYTD 60
OY 61 GDVKLTQSMATIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRVAYSKEDETLKV 120
DB 61 GDVKLTQSMATIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRVAYSKEDETLKV 120
OY 121 DFLSKLPKMKFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCKLDAFPKLVCFK 180
DB 121 DFLSKLPKMKFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCKLDAFPKLVCFK 180
OY 121 DFLSKLPKMKFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCKLDAFPKLVCFK 180
DB 121 DFLSKLPKMKFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCKLDAFPKLVCFK 180
OY 181 KRIEAIPQIDKYLKSSKTYAMPLOGWQATFGGDDHPK 218
DB 181 KRIEAIPQIDKYLKSSKTYAMPLOGWQATFGGDDHPK 218

RESULT 5
AAR72793
ID AAR72793 standard; Protein: 232 AA.
XX
AC AAR72793;
XX
DT 16-OCT-1995 (first entry)
XX
DE Glutathione-s-transferase.
XX
KW SDR-1; senescent cell-derived inhibitor; DNA synthesis;
KW senescence; cell proliferation; cancer; therapeutic; vulnery;
KW fusion protein; glutathione-s-transferase.
XX
OS Schistosoma japonicum.
XX
PN WO9506415-A.
XX
PD 09-MAR-1995.
XX
PF 26-AUG-1994; 94WO-US09700.
XX
PR 13-JUL-1994; 94US-0274535.
PR 30-AUG-1993; 93US-0113372.
PR 17-NOV-1993; 93US-0153564.
PR 03-JAN-1994; 94US-0160814.
PR 25-FEB-1994; 94US-0203535.
PR 15-APR-1994; 94US-0228420.
PR 30-JUN-1994; 94US-0268439.
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Smith JR;
XX
WPI: 1995-131101/17.
XX
DR N-PSDB; AAO86774.
XX
PT Nucleic acid encoding a protein or polypeptide that inhibits DNA
PS synthesis in a recipient cell - useful to inhibit cell
```

PT proliferation in tumour cells, treat wound or burn tissue, or as
 PT an antiviral or antiparasitic agent
 XX
 PS Disclosure: Page 130; 169pp; English.
 XX
 CC The senescent cell-derived inhibitor-1 (SDI-1) inhibits DNA
 CC synthesis in a recipient cell and is capable of associating with a
 CC cyclin (esp. cyclin D1) or cyclin-dependent kinase (esp. CDK2).
 CC Production of a fusion protein of SDI-1 and glutathione-S-transferase
 CC (AAR77793) facilitated the preparation and isolation of recombinant
 CC SDI-1.
 CC
 XX
 SQ Sequence 232 AA;
 Query Match 98.7%; Score 1149; DB 16; Length 232;
 Best Local Similarity 98.6%; Pred. No. 7.2e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGKINGLVOPTRLLLEYLEEKEEHLERDEDEKRNKKFELGLEPNNLPYYID 60
 DB 1 MSPILGKINGLVOPTRLLLEYLEEKEEHLERDEDEKRNKKFELGLEPNNLPYYID 60
 QY 61 GDVKLTQSMALIRYADKHNMLGGSPKERAELSMLEGAVIDIRYVSRIAYSKDEETLKV 120
 DB 61 GDVKLTQSMALIRYADKHNMLGGSPKERAELSMLEGAVIDIRYVSRIAYSKDEETLKV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPPLVSFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPPLVSFK 180
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMATFGGDHPPK 218
 DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMATFGGDHPPK 218
 RESULT 6
 AAW73909
 ID AAW73909 standard; Protein: 240 AA.
 XX
 AC AAW73909;
 XX
 DT 13-APR-1999 (first entry)
 XX
 DE GST protein sequence.
 XX
 KW Cell surface nonexpressive functional protein; cell-permeating protein;
 KW protein screening; protein production; GST; glutathione-S-transferase.
 XX
 OS Schistosoma japonicum.
 XX
 PN JPI1009278-A.
 XX
 PD 19-JAN-1999.
 XX
 PF 23-JUN-1997; 97JP-0165788.
 XX
 PR 23-JUN-1997; 97JP-0165788.
 XX
 PA (SUME) SUMITOMO ELECTRIC IND CO.
 XX
 DR WPI: 1999-145895/13.
 DR N-PSDB: AAX01312.
 XX
 PT Microbe having a nonexpressive functional protein on its surface -
 PT for screening and large-scale preparation of a selected function
 PT protein
 XX
 XX Disclosure: Page 8-9; 13pp; Japanese.
 CC This sequence is the glutathione-S-transferase (GST) protein, which
 CC was expressed in the microbe of the invention. The microbe has a cell
 CC surface nonexpressive functional protein on its surface, fused to a
 CC support protein derived from a cell-permeating protein. The microbe can

CC be used for screening for biological activity of a cell surface
 CC nonexpressive protein. It can also be used for preparation of a selected
 CC functional protein in a large amounts.
 CC
 XX
 SQ Sequence 240 AA;
 Query Match 98.7%; Score 1149; DB 20; Length 240;
 Best Local Similarity 98.6%; Pred. No. 7.6e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGKINGLVOPTRLLLEYLEEKEEHLERDEDEKRNKKFELGLEPNNLPYYID 60
 DB 1 MSPILGKINGLVOPTRLLLEYLEEKEEHLERDEDEKRNKKFELGLEPNNLPYYID 60
 QY 61 GDVKLTQSMALIRYADKHNMLGGSPKERAELSMLEGAVIDIRYVSRIAYSKDEETLKV 120
 DB 61 GDVKLTQSMALIRYADKHNMLGGSPKERAELSMLEGAVIDIRYVSRIAYSKDEETLKV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPPLVSFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPPLVSFK 180
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMATFGGDHPPK 218
 DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMATFGGDHPPK 218
 RESULT 7
 AAB31612
 ID AAB31612 standard; Protein: 245 AA.
 XX
 AC AAB31612;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Amino acid sequence of glutathion-S-transferase (GST).
 XX
 KW Heat shock protein; Hsp: Th1 response; Th1 cell; CD4+ T lymphocyte cell;
 KW lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;
 KW glutathion-S-transferase; GST.
 XX
 OS Unidentified.
 XX
 PN WO200104344-A2.
 XX
 PD 18-JAN-2001.
 XX
 PF 10-JUL-2000; 2000WO-US18828.
 XX
 PR 08-JUL-1999; 99US-0143757.
 XX
 PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
 XX
 PI Siegel M, Chu NR, Mizzzen LA;
 XX
 DR WPI: 2001-138361/14.
 DR N-PSDB: AAF25015.
 XX
 PT Screening for compounds that stimulate Th1-like responses in CD4+ T
 PT lymphocyte cells -
 XX
 XX Example 9; Fig 8; 88pp; English.
 CC The present sequence represents a glutathion-S-transferase (GST), which
 CC was used linked to HPV16 E7 protein, for purification purposes. The E7
 CC protein was used to construct a fusion protein with Mycobacterium bovis
 CC heat shock proteins (Hsp). The fusion proteins are used in the method of
 CC the invention. The specification describes a method of determining
 CC whether a compound stimulates a Th1-like response. Th1 cells are a subset
 CC of CD4+ T lymphocyte cells. The method comprises contacting naive
 CC lymphocytes in vitro with a fusion protein comprising at least a fragment
 CC of Hsp, and then detecting the Th1-like response exhibited by the cell
 CC sample. The proteins which may be used in the method of the invention are

CC Hsp65, Hsp40, Hsp10, Hsp60, and Hsp71. The method may be used to identify
 CC compounds that stimulate Th1-like responses in response to microbial
 CC pathogens.
 CC
 XX
 SQ Sequence 245 AA;
 Query Match 98.7%; Score 1149; DB 22; Length 245;
 Best Local Similarity 98.6%; Pred. No. 7.9e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGEPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGEPNLPYYID 60
 QY 61 GDVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRIVASKDFETLKV 120
 DB 61 GDVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRIVASKDFETLKV 120
 QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHDPFMYLDALDVLYMDPCLDAFRLVCFK 180
 DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDALDVLYMDPCLDAFRLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKTIAMPLOGMQATFGGDPHPK 218
 DB 181 KRIEAIPOIDKYLKSSKTIAMPLOGMQATFGGDPHPK 218
 RESULT 8
 AAM74114
 ID AAM74114 standard; peptide: 247 AA.
 XX
 AC AAM74114;
 XX
 DT 04-MAY-1999 (first entry)
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein: hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JF;
 PI Mahony DJ, Patterson CA, Singleton J;
 DR WPI; 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimerae and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example; Page 217; 294pp; English.
 XX
 CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),

CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 XX
 SQ Sequence 247 AA;
 Query Match 98.7%; Score 1149; DB 20; Length 247;
 Best Local Similarity 98.6%; Pred. No. 7.9e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGEPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGEPNLPYYID 60
 QY 61 GDVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRIVASKDFETLKV 120
 DB 61 GDVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRIVASKDFETLKV 120
 QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHDPFMYLDALDVLYMDPCLDAFRLVCFK 180
 DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDALDVLYMDPCLDAFRLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKTIAMPLOGMQATFGGDPHPK 218
 DB 181 KRIEAIPOIDKYLKSSKTIAMPLOGMQATFGGDPHPK 218
 RESULT 9
 AAM74120
 ID AAM74120 standard; peptide: 247 AA.
 XX
 AC AAM74120;
 XX
 DT 04-MAY-1999 (first entry)
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein: hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JF;
 PI Mahony DJ, Patterson CA, Singleton J;
 DR WPI; 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the

PT gastro-intestinal tract and related nucleic acid - chimaeras and
PT antibodies, used to deliver therapeutic or diagnostic agents to, or
XX through, the gastrointestinal tract, e.g. insulin or leuprolide
PS Example; Page 221; 294pp; English.

CC This sequence represents a fusion protein between
CC glutathione-S-transferase and a gastro-intestinal transport protein
CC binding peptide. The invention relates to purified proteins (I) that bind
CC specifically to at least one of the gastro-intestinal (GI) tract
CC receptors human intestinal peptide-associated transporter (HPT1),
CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
CC active transport of therapeutic agents through human and animal GI tissue
CC (into the blood) for in vivo delivery, particularly for treatment or
CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
CC cancer, migraine, or angina pectoris. Specifically they are used to
CC deliver insulin or leuprolide, but many other suitable therapeutic agents
CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
CC and antigens. (I) may also provide targeting to the GI tract. Other uses
CC of (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunosays. The antibodies are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.

SO Sequence 247 AA;

Query Match 98.7%; Score 1149; DB 20; Length 247;
Best Local Similarity 98.6%; Pred. No. 7.9e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYKINGVLQPTLLLEYLEEKYEHLHYERDEGDKNNKKFELGLEPNNLPYYID 60
DB 1 MSPILGYKINGVLQPTLLLEYLEEKYEHLHYERDEGDKNNKKFELGLEPNNLPYYID 60

QY 61 GDVKLTQSMALIRYADKHNMLGSGPKERAETSMLEGAVIDIRGVSRIVASKDEFETLKV 120

DB 61 GDVKLTQSMALIRYADKHNMLGSGPKERAETSMLEGAVIDIRGVSRIVASKDEFETLKV 120

QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHDPFMYDALDVVLYMDPCLDAFPPLVSFK 180

DB 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHDPFMYDALDVVLYMDPCLDAFPPLVSFK 180

QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGSDHPPK 218

DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGSDHPPK 218

RESULT 10

ID AAM74103 standard; peptide: 247 AA.

AC AAM74103;

XX 04-MAY-1999 (first entry)

DE GST/GI transport receptor binding protein fusion sequence.

KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

XX MO9851325-A2.

XX 19-NOV-1998.

PF 15-MAY-1998; 98WO-US10088.

PR 15-MAY-1997; 97US-0046595.
XX (CYTO-) CYTOGEN CORP.
PA (ELAN-) ELAN CORP PLC.
XX
PI Alvarez VL, Bellinka BA, Cagney GM, Carter JM, Lambkin J;
PI Mahony DJ, Patterson CA, Singleton J;
XX WPI: 1999-009568/01.

XX New proteins that bind specifically to receptors in the
PT gastro-intestinal tract and related nucleic acid - chimaeras and
PT antibodies, used to deliver therapeutic or diagnostic agents to, or
XX through, the gastrointestinal tract, e.g. insulin or leuprolide
PS Example; Page 209; 294pp; English.

CC This sequence represents a fusion protein between
CC glutathione-S-transferase and a gastro-intestinal transport protein
CC binding peptide. The invention relates to purified proteins (I) that bind
CC specifically to at least one of the gastro-intestinal (GI) tract
CC receptors human intestinal peptide-associated transporter (HPT1),
CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
CC active transport of therapeutic agents through human and animal GI tissue
CC (into the blood) for in vivo delivery, particularly for treatment or
CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
CC cancer, migraine, or angina pectoris. Specifically they are used to
CC deliver insulin or leuprolide, but many other suitable therapeutic agents
CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
CC and antigens. (I) may also provide targeting to the GI tract. Other uses
CC of (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunosays. The antibodies are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.

SO Sequence 247 AA;

Query Match 98.7%; Score 1149; DB 20; Length 247;
Best Local Similarity 98.6%; Pred. No. 7.9e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYKINGVLQPTLLLEYLEEKYEHLHYERDEGDKNNKKFELGLEPNNLPYYID 60

DB 1 MSPILGYKINGVLQPTLLLEYLEEKYEHLHYERDEGDKNNKKFELGLEPNNLPYYID 60

QY 61 GDVKLTQSMALIRYADKHNMLGSGPKERAETSMLEGAVIDIRGVSRIVASKDEFETLKV 120

DB 61 GDVKLTQSMALIRYADKHNMLGSGPKERAETSMLEGAVIDIRGVSRIVASKDEFETLKV 120

QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHDPFMYDALDVVLYMDPCLDAFPPLVSFK 180

DB 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHDPFMYDALDVVLYMDPCLDAFPPLVSFK 180

QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGSDHPPK 218

DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGSDHPPK 218

RESULT 11

ID AAM74112 standard; peptide: 248 AA.

AC AAM74112;

XX 04-MAY-1999 (first entry)

DE GST/GI transport receptor binding protein fusion sequence.

KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;

KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.
 XX Homo sapiens.
 XX W09851325-A2.
 PN 19-NOV-1998.
 XX 15-MAY-1998; 98WO-US10088.
 PD 15-MAY-1997; 97US-0046595.
 PR (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin LJ;
 PI Omahony DJ, Patterson CA, Singleton J;
 DR WPI: 1999-009568/01.
 XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example; Page 215-216; 294pp; English.
 XX This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPIT),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hs1). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX Sequence 248 AA;
 SQ
 Query Match 98.7%; Score 1149; DB 20; Length 248;
 Best Local Similarity 98.6%; Pred. No. 8e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ID AAW74113 standard; peptide: 248 AA.
 XX AC AAW74113;
 XX 04-MAY-1999 (first entry)
 DT GST/GI transport receptor binding protein fusion sequence.
 DE
 XX Gastro-intestinal transport receptor; binding protein; hSi; hPIT1;
 KW D2H; hPPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.
 XX
 XX Homo sapiens.
 OS
 XX W09851325-A2.
 PN 19-NOV-1998.
 PD 15-MAY-1998; 98WO-US10088.
 PF 15-MAY-1997; 97US-0046595.
 PR (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin LJ;
 PI Omahony DJ, Patterson CA, Singleton J;
 DR WPI: 1999-009568/01.
 XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example; Page 216; 294pp; English.
 XX This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPIT),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hs1). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX Sequence 248 AA;
 SQ
 Query Match 98.7%; Score 1149; DB 20; Length 248;
 Best Local Similarity 98.6%; Pred. No. 8e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTVHPDFMLDALDVLYMDPMLDAFPKLVSRK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLDALDVLYMDPMLDAFPKLVSRK 180
 QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGQWATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGQWATFGGSDHPK 218

RESULT 13

AAW74121 ID AAW74121 standard; peptide: 248 AA.

AAW74121; AC

04-MAY-1999 (first entry)

GST/GI transport receptor binding protein fusion sequence.

KW Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

PN WO9851325-A2.

PD 19-NOV-1998.

PF 15-MAY-1998; 98WO-US10088.

PR 15-MAY-1997; 97US-0046595.

PA (CYTO-) CYTOGEN CORP.

PA (ELAN-) ELAN CORP PLC.

PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JJ;
 PI Omanony DJ, Patterson CA, Singleton J;

DR WPI; 1999-009568/01.

PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example; Page 222; 294pp; English.

CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hsi). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (1) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (11) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

CC Sequence 248 AA;

Query Match 98.7%, Score 1149; DB 20; Length 248;

Best Local Similarity 98.6%; Pred. No. 8e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLHYERDEGGRANKKFEGLLEPPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLHYERDEGGRANKKFEGLLEPPNLPYYID 60
 QY 61 GGVKLTQSMATIRYIADKNNMLGSGPKERAEITSMLGAVLDIRYGSRIYASKDEFTLV 120
 DB 61 GGVKLTQSMATIRYIADKNNMLGSGPKERAEITSMLGAVLDIRYGSRIYASKDEFTLV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTVHPDFMLDALDVLYMDPMLDAFPKLVSRK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLDALDVLYMDPMLDAFPKLVSRK 180
 QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGQWATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGQWATFGGSDHPK 218

RESULT 14

AAW74122 ID AAW74122 standard; peptide: 248 AA.

AAW74122; AC

04-MAY-1999 (first entry)

GST/GI transport receptor binding protein fusion sequence.

KW Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

PN WO9851325-A2.

PD 19-NOV-1998.

PF 15-MAY-1998; 98WO-US10088.

PR 15-MAY-1997; 97US-0046595.

PA (CYTO-) CYTOGEN CORP.

PA (ELAN-) ELAN CORP PLC.

PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JJ;
 PI Omanony DJ, Patterson CA, Singleton J;

DR WPI; 1999-009568/01.

PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example; Page 222-223; 294pp; English.

CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hsi). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents

CC and antigens. (1) may also provide targeting to the GI tract. Other uses
CC of (1) are: (1) to determine the level of specified receptors in a sample
CC (in a binding assay); and (11) to screen for molecules that bind (1).
CC Immunogenic analogues or derivatives of (1) are used to raise antibodies
CC and in immunoassays. The antibodies are used to locate, detect and
CC measure (1), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
XX
SQ Sequence 248 AA:
Query Match 98.7%; Score 1149; DB 20; Length 248;
Best Local Similarity 98.6%; Pred. No. 8e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMKNKFEGLPEPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMKNKFEGLPEPNLPYYID 60
QY 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVLDIRYGSRIAYSKDEFETLKV 120
DB 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVLDIRYGSRIAYSKDEFETLKV 120
QY 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLVYMDPKLDAFPKLVYSK 180
DB 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLVYMDPKLDAFPKLVYSK 180
QY 181 KRIEAIPOIDKYLKSSKTYAMPLOGWQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYAMPLOGWQATFGGDDHPK 218
RESULT 15
AAW74100
ID AAW74100 standard; peptide: 248 AA.
XX
AC AAW74100;
XX
DT 04-MAY-1999 (first entry)
XX
DE GST/GI transport receptor binding protein fusion sequence.
XX
KW Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
KW D2H; hhep1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW therapeutic agent delivery; therapy; fusion protein.
XX
OS Homo sapiens.
XX
PN WO9851325-A2.
XX
PD 19-NOV-1998.
XX
PF 15-MAY-1998; 98WO-US10088.
XX
PR 15-MAY-1997; 97US-0046595.
XX
PA (CYTO-) CYTOGEN CORP.
PA (ELAN-) ELAN CORP PLC.
XX
PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambdin TJ;
PI Omahony DJ, Patterson CA, Singleton J;
XX
DR WPI, 1999-009568/01.
XX
PT New proteins that bind specifically to receptors in the
PT gastro-intestinal tract and related nucleic acid - chimaeras and
PT antibodies, used to deliver therapeutic or diagnostic agents to, or
PT through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
PS Example: Page 207; 294pp; English.
XX
CC This sequence represents a fusion protein between

CC glutathione-S-transferase and a gastro-intestinal transport protein
CC binding peptide. The invention relates to purified proteins (1) that bind
CC specifically to at least one of the gastro-intestinal (GI) tract
CC receptors human intestinal peptide-associated transporter (HPT1),
CC hhep1, D2H and human sucrose-isomaltase complex (hsi). (1) provide
CC active transport of therapeutic agents through human and animal GI tissue
CC (into the blood) for in vivo delivery, particularly for treatment or
CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
CC cancer, migraine, or angina pectoris. Specifically they are used to
CC deliver insulin or leuprolide, but many other suitable therapeutic agents
CC are disclosed. Including genes or inhibitory nucleic acid, imaging agents
CC and antigens. (1) may also provide targeting to the GI tract. Other uses
CC of (1) are: (1) to determine the level of specified receptors in a sample
CC (in a binding assay); and (11) to screen for molecules that bind (1)
CC Immunogenic analogues or derivatives of (1) are used to raise antibodies
CC and in immunoassays. The antibodies are used to locate, detect and
CC measure (1), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
XX
SQ Sequence 248 AA:
Query Match 98.7%; Score 1149; DB 20; Length 248;
Best Local Similarity 98.6%; Pred. No. 8e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMKNKFEGLPEPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMKNKFEGLPEPNLPYYID 60
QY 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVLDIRYGSRIAYSKDEFETLKV 120
DB 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVLDIRYGSRIAYSKDEFETLKV 120
QY 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLVYMDPKLDAFPKLVYSK 180
DB 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLVYMDPKLDAFPKLVYSK 180
QY 181 KRIEAIPOIDKYLKSSKTYAMPLOGWQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYAMPLOGWQATFGGDDHPK 218
Search completed: May 20, 2003, 11:33:55
Job time : 19.4058 secs

Tue May 20 13:19:21 2003

us-10-081-408-5.ra1

Page 1

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 7.02587 Seconds
(without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-5
Perfect score: 1164
Sequence: 1 MSPILGWKIKGLVQPTRL.....IAMPLOGMAGTFCGDPHPK 218

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	98.7	218	1 US-08-470-837-26	Sequence 26, Appl
2	1149	98.7	218	4 US-08-868-452-26	Sequence 26, Appl
3	1149	98.7	231	4 US-08-129-722A-2	Sequence 2, Appl
4	1149	98.7	232	4 US-08-327-874A-6	Sequence 6, Appl
5	1149	98.7	232	5 PCT-US94-09700-6	Sequence 6, Appl
6	1149	98.7	272	4 US-08-910-820-4	Sequence 4, Appl
7	1149	98.7	272	4 US-08-910-820-6	Sequence 4, Appl
8	1149	98.7	282	4 US-08-910-820-3	Sequence 3, Appl
9	1149	98.7	282	4 US-08-910-820-5	Sequence 3, Appl
10	1149	98.7	307	2 US-08-216-894-6	Sequence 6, Appl
11	1149	98.7	307	4 US-09-115-746-6	Sequence 6, Appl
12	1149	98.7	331	4 US-09-217-228-6	Sequence 6, Appl
13	1149	98.7	352	1 US-08-395-507-1	Sequence 1, Appl
14	1149	98.7	354	2 US-08-216-894-4	Sequence 4, Appl
15	1149	98.7	354	4 US-09-115-746-4	Sequence 4, Appl
16	1149	98.7	362	1 US-08-395-507-2	Sequence 2, Appl
17	1149	98.7	397	4 US-08-327-874A-11	Sequence 11, Appl
18	1149	98.7	397	5 PCT-US94-09700-11	Sequence 11, Appl
19	1149	98.7	412	4 US-09-366-009-34	Sequence 34, Appl
20	1149	98.7	422	4 US-09-217-228-7	Sequence 7, Appl
21	1149	98.7	426	3 US-08-737-248-4	Sequence 4, Appl
22	1149	98.7	435	5 PCT-US95-04439-1	Sequence 1, Appl
23	1149	98.7	439	4 US-08-506-296B-67	Sequence 67, Appl
24	1149	98.7	442	4 US-08-506-296B-70	Sequence 70, Appl
25	1149	98.7	443	4 US-08-506-296B-76	Sequence 76, Appl
26	1149	98.7	447	4 US-08-506-296B-73	Sequence 73, Appl
27	1149	98.7	472	2 US-08-216-894-10	Sequence 10, Appl

28	1149	98.7	472	4 US-09-115-746-10	Sequence 10, Appl
29	1149	98.7	514	4 US-08-974-549A-605	Sequence 605, App
30	1149	98.7	515	4 US-08-974-549A-604	Sequence 604, App
31	1149	98.7	517	4 US-08-974-549A-606	Sequence 606, App
32	1149	98.7	530	4 US-08-974-549A-603	Sequence 603, App
33	1149	98.7	536	3 US-08-974-549A-602	Sequence 602, App
34	1149	98.7	542	4 US-08-506-296B-69	Sequence 69, Appl
35	1149	98.7	545	4 US-08-506-296B-75	Sequence 75, Appl
36	1149	98.7	547	4 US-08-506-296B-72	Sequence 72, Appl
37	1149	98.7	547	4 US-08-506-296B-66	Sequence 66, Appl
38	1149	98.7	559	4 US-08-216-894-2	Sequence 2, Appl
39	1149	98.7	564	2 US-09-115-746-2	Sequence 2, Appl
40	1149	98.7	579	2 US-08-864-224-11	Sequence 11, Appl
41	1149	98.7	632	4 US-08-506-296B-74	Sequence 74, Appl
42	1149	98.7	635	4 US-08-506-296B-71	Sequence 71, Appl
43	1149	98.7	643	2 US-08-216-894-8	Sequence 8, Appl
44	1149	98.7	643	2 US-09-115-746-8	Sequence 8, Appl
45	1149	98.7	643	4 US-09-115-746-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-470-837-26
Sequence 26, Application US/08470837
Patent No. 5800811
GENERAL INFORMATION:
APPLICANT: Nimmi, Marcel E.
APPLICANT: Hall, Frederick L.
APPLICANT: Tuan, Tai-Lan
APPLICANT: Wu, Lingtao
APPLICANT: Cheung, David T.
TITLE OF INVENTION: Transforming Growth Factor B Fusion
TITLE OF INVENTION: and
TITLE OF INVENTION: Their Use in Wound Healing
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESSES:
ADDRESS: Merchant & Gould
STREET: 11150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025-3395
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,837
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sharp, Janice A.
REGISTRATION NUMBER: 34,051
REFERENCE/DOCKET NUMBER: 30630-1US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-837-26
Query Match 98.7% Score 1149: DB 1: Length 218:
Best Local Similarity 98.6% Pred. NO. 6.8e-114:
Matches 215: Conservative 0: Mismatches 3: Indels 0: Gaps 0:
Oy 1 MSPILGWKIKGLVQPTRLLEYLEKEEHLERDECDKRNKKFEELGFPNLPYYID 60

Db 1 MSPIIGYWKIKGLVOPTRLLLEYLEKEEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Qy 61 GDVKTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVIDIRYGVSRIVASKEDETLKV 120
Db 61 GDVKTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVIDIRYGVSRIVASKEDETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Db 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Qy 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPK 218
Db 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPK 218

RESULT 2

US-08-868-452-26
; Sequence 26, Application US/08868452C
; Patent No. 6352972
; GENERAL INFORMATION:
; APPLICANT: Marcel E. Nimmi
; APPLICANT: Frederick L. Hall
; APPLICANT: Lingtao Wu
; APPLICANT: Bo Han
; APPLICANT: Edwin Shors
; TITLE OF INVENTION: BONE MORPHOGENETIC PROTEINS AND THEIR
; TITLE OF INVENTION: USE IN BONE GROWTH
; FILE REFERENCE: 17972-11
; CURRENT APPLICATION NUMBER: US/08/868,452C
; CURRENT FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-868-452-26

Query Match 98.7%; Score 1149; DB 4; Length 218;
Best Local Similarity 98.6%; Pred. No. 6,8e-114;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MSPIIGYWKIKGLVOPTRLLLEYLEKEEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Db 1 MSPIIGYWKIKGLVOPTRLLLEYLEKEEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Qy 61 GDVKTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVIDIRYGVSRIVASKEDETLKV 120
Db 61 GDVKTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVIDIRYGVSRIVASKEDETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Db 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Qy 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPK 218
Db 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPK 218

RESULT 3

US-08-129-722A-2
; Sequence 2, Application US/08129722A
; Patent No. 6303369
; GENERAL INFORMATION:
; APPLICANT: Spana, Carl
; APPLICANT: Fargnoli, Joseph
; APPLICANT: Bolen, Joseph B.
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000

CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/129,722A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC25
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 231 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-129-722A-2

Query Match 98.7%; Score 1149; DB 4; Length 231;
Best Local Similarity 98.6%; Pred. No. 7,4e-114;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MSPIIGYWKIKGLVOPTRLLLEYLEKEEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Db 1 MSPIIGYWKIKGLVOPTRLLLEYLEKEEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Qy 61 GDVKTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVIDIRYGVSRIVASKEDETLKV 120
Db 61 GDVKTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVIDIRYGVSRIVASKEDETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Db 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Qy 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPK 218
Db 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPK 218

RESULT 4

US-08-327-874A-6
; Sequence 6, Application US/08327874A
; Patent No. 6372249
; GENERAL INFORMATION:
; APPLICANT: BAYLOR COLLEGE OF MEDICINE
; APPLICANT: SMITH, JAMES R.
; APPLICANT: DRUTZ, DAVID J.
; APPLICANT: WILSON, DEBORAH R.
; APPLICANT: ZUMSTEIN, LOUIS A.
; TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
; TITLE OF INVENTION: DNA SYNTHESIS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS & WELLS
; STREET: 200 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10166
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,874A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT US94/09700
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/327,874
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: NORTON, GERARD P.
REGISTRATION NUMBER: 36,621
REFERENCE/DOCKET NUMBER: 3634-8-CIP10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 878-3148
TELEFAX: (212) 878-8375
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Schistosoma japonicum
IMMEDIATE SOURCE:
CLONE: GST
US-08-327-874A-6

Query Match 98.7%; Score 1149; DB 4; Length 232;
Best Local Similarity 98.6%; Pred. No. 7.4e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYKIKGLVQPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEPNNLPYYID 60
DB 1 MSPILGYKIKGLVQPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEPNNLPYYID 60
QY 61 GDVLTOSMAIIRYADRNHMLGSPKRAEISMLEGAVLDIRYVSRIAYSKDEETLV 120
DB 61 GDVLTOSMAIIRYADRNHMLGSPKRAEISMLEGAVLDIRYVSRIAYSKDEETLV 120
QY 121 DFLSLPMLKMFEDRLSHKTYTLNGDHTHPDMLYDALDVLVYNDPCLDAFPLVSFK 180
DB 121 DFLSLPMLKMFEDRLSHKTYTLNGDHTHPDMLYDALDVLVYNDPCLDAFPLVSFK 180
QY 161 KRIEAIPOIDKYLKSSXYIAMPLOGWQATFGGDDHPK 218
DB 161 KRIEAIPOIDKYLKSSXYIAMPLOGWQATFGGDDHPK 218

RESULT 5
PCT-US94-09700-6
Sequence 6, Application PC/TUS9409700
GENERAL INFORMATION:
APPLICANT: BAYLOR COLLEGE OF MEDICINE
APPLICANT: SMITH, JAMES R.
TITLE OF INVENTION: SENESECENT CELL DERIVED INHIBITORS OF
TITLE OF INVENTION: DNA SYNTHESIS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWREY & SIMON
STREET: 1299 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09700
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: AUERBACH, JEFFREY I.
REGISTRATION NUMBER: 32,680
REFERENCE/DOCKET NUMBER: 225-102-CIP7-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 383-7451
TELEFAX: (202) 383-6610
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Schistosoma japonicum
IMMEDIATE SOURCE:
CLONE: GST
PCT-US94-09700-6

Query Match 98.7%; Score 1149; DB 5; Length 232;
Best Local Similarity 98.6%; Pred. No. 7.4e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYKIKGLVQPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEPNNLPYYID 60
DB 1 MSPILGYKIKGLVQPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEPNNLPYYID 60

QY 61 GGVKLTQSMATIRYIADHNHMGSPKERAETSMLEGAVIDIRYGVSRVAYSKDFETLKV 120
DB 61 GGVKLTQSMATIRYIADHNHMGSPKERAETSMLEGAVIDIRYGVSRVAYSKDFETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPMLDAFPKLVYFK 180
DB 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPMLDAFPKLVYFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOWATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOWATFGGDDHPK 218

RESULT 6

US-08-910-820-4
Sequence 4, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gfan
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/910,820
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-4

Query Match 98.7%; Score 1149; DB 4; Length 272;
Best Local Similarity 98.6%; Pred. No. 9.3e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGVKIKGLVOPTRLLEYLEEKYEENHYERDEGDKMNKKEFELGEPNLPYYID 60
DB 1 MSPILGVKIKGLVOPTRLLEYLEEKYEENHYERDEGDKMNKKEFELGEPNLPYYID 60
QY 61 GGVKLTQSMATIRYIADHNHMGSPKERAETSMLEGAVIDIRYGVSRVAYSKDFETLKV 120
DB 61 GGVKLTQSMATIRYIADHNHMGSPKERAETSMLEGAVIDIRYGVSRVAYSKDFETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPMLDAFPKLVYFK 180
DB 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPMLDAFPKLVYFK 180

QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOWATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOWATFGGDDHPK 218

RESULT 7

US-08-910-820-6
Sequence 6, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gfan
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/910,820
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-6

Query Match 98.7%; Score 1149; DB 4; Length 272;
Best Local Similarity 98.6%; Pred. No. 9.3e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGVKIKGLVOPTRLLEYLEEKYEENHYERDEGDKMNKKEFELGEPNLPYYID 60
DB 1 MSPILGVKIKGLVOPTRLLEYLEEKYEENHYERDEGDKMNKKEFELGEPNLPYYID 60
QY 61 GGVKLTQSMATIRYIADHNHMGSPKERAETSMLEGAVIDIRYGVSRVAYSKDFETLKV 120
DB 61 GGVKLTQSMATIRYIADHNHMGSPKERAETSMLEGAVIDIRYGVSRVAYSKDFETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPMLDAFPKLVYFK 180
DB 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPMLDAFPKLVYFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOWATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOWATFGGDDHPK 218

RESULT 8
US-08-910-820-3

Sequence 3, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-3

Query Match 98.7%; Score 1149; DB 4; Length 282;
Best Local Similarity 98.6%; Pred. No. 9.8e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWYKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKWNNKFEELGLEFPNLPYYID 60
DB 1 MSPILGWYKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKWNNKFEELGLEFPNLPYYID 60
QY 61 GGVKLTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGSRIAYSKDEETLKV 120
DB 61 GGVKLTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGSRIAYSKDEETLKV 120
QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPKCLDAFPKLYVSK 180
DB 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPKCLDAFPKLYVSK 180
QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPKCLDAFPKLYVSK 180
DB 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPKCLDAFPKLYVSK 180
QY 181 KRIEAIPOIDKYLKSSKXIYAMPLOGWQATFGGSDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXIYAMPLOGWQATFGGSDHPK 218

RESULT 9
US-08-910-820-5
Sequence 5, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-5

Query Match 98.7%; Score 1149; DB 4; Length 282;
Best Local Similarity 98.6%; Pred. No. 9.8e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWYKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKWNNKFEELGLEFPNLPYYID 60
DB 1 MSPILGWYKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKWNNKFEELGLEFPNLPYYID 60
QY 61 GGVKLTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGSRIAYSKDEETLKV 120
DB 61 GGVKLTQSMALIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGSRIAYSKDEETLKV 120
QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPKCLDAFPKLYVSK 180
DB 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPKCLDAFPKLYVSK 180
QY 181 KRIEAIPOIDKYLKSSKXIYAMPLOGWQATFGGSDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXIYAMPLOGWQATFGGSDHPK 218

RESULT 10
US-08-216-894-6
Sequence 6, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216.894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-6

Query Match 98.7%; Score 1149; DB 2; Length 307;
Best Local Similarity 98.6%; Pred. No. 1.1e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGYKIKGLVPTLLLEYLEEKYEHLHYEDGDKMKNKFEGLFEPNLPYYID 60
DB 1 MSPILGYKIKGLVPTLLLEYLEEKYEHLHYEDGDKMKNKFEGLFEPNLPYYID 60
OY 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
DB 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
OY 121 DFLSKLPEMLKFEEDRLCHKTYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVSEK 180
DB 121 DFLSKLPEMLKFEEDRLCHKTYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVSEK 180
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGGDHPK 218

RESULT 11

US-09-115-746-6
Sequence 6, Application US/09115746
Patent No. 6228601
GENERAL INFORMATION:
APPLICANT: Kirchoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAPOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115.746
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216.894
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-6

Query Match 98.7%; Score 1149; DB 4; Length 307;
Best Local Similarity 98.6%; Pred. No. 1.1e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGYKIKGLVPTLLLEYLEEKYEHLHYEDGDKMKNKFEGLFEPNLPYYID 60
DB 1 MSPILGYKIKGLVPTLLLEYLEEKYEHLHYEDGDKMKNKFEGLFEPNLPYYID 60
OY 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
DB 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
OY 121 DFLSKLPEMLKFEEDRLCHKTYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVSEK 180
DB 121 DFLSKLPEMLKFEEDRLCHKTYLNGDHYTHDPFMLYDALDVLVYMDPCLDAFPKLVSEK 180
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGGDHPK 218

RESULT 12

US-09-217-228-6
Sequence 6, Application US/09217228
Patent No. 6323178
GENERAL INFORMATION:
APPLICANT: Butler, Jon P.
APPLICANT: Hale, John E.
APPLICANT: Heath Jr., William F.
APPLICANT: Schoner, Brigitte E.
APPLICANT: Helman, Mark L.
APPLICANT: Becker, Gerald W.
APPLICANT: Varshavsky, Alexander D.
TITLE OF INVENTION: Beta-lipotropin and Uses Thereof
FILE REFERENCE: X-12139
CURRENT FILING DATE: 1998-12-21
CURRENT APPLICATION NUMBER: US/09/217.228
NUMBER OF SEQ ID NOS: 35
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 6
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GST fusion
US-09-217-228-6

Query Match 98.7%; Score 1149; DB 4; Length 331;
Best Local Similarity 98.6%; Pred. No. 1.2e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGYKIKGLVPTLLLEYLEEKYEHLHYEDGDKMKNKFEGLFEPNLPYYID 60
DB 1 MSPILGYKIKGLVPTLLLEYLEEKYEHLHYEDGDKMKNKFEGLFEPNLPYYID 60
OY 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
DB 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120

OY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVVLYMDPMLDAFRLVSEK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVVLYMDPMLDAFRLVSEK 180
OY 181 KRIEAIPOIDKYLKSKSIYAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSKSIYAMPLOGWQATFGGDHPK 218

RESULT 13

US-08-395-507-1
Sequence 1, Application US/08395507
Patent No. 5578456
GENERAL INFORMATION:
APPLICANT: Fujimura, Katsuya
APPLICANT: Ueno, Eiichi
APPLICANT: Fujii, No. 5578456uyuk1
APPLICANT: Okada, Masahisa
TITLE OF INVENTION: Anti-tetraponema Pallidum Antibody
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pollock, Vande Sande & Priddy
STREET: 1990 M Street, NW, Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,507
FILING DATE: 27-FEB-1995
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP Hei-6-54672
FILING DATE: 28-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Amernick, Burton A.
REGISTRATION NUMBER: 24,852
REFERENCE/DOCKET NUMBER: 1581/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 331-7111
TELEFAX: (202) 293-6229
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: rabbit
FEATURE:
NAME/KEY: Protein
LOCATION: 1..352
OTHER INFORMATION: /note= "G15 Antigen"
US-08-395-507-1

Query Match 98.7%; Score 1149; DB 1; Length 352;
Best Local Similarity 98.6%; Pred. No. 1.3e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLHYERDEGDKWNNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLHYERDEGDKWNNKKFELGLEFPNLPYYID 60
OY 61 GDVKLTQSMATIRYIADKHNMLGSGPKRAEISMLEGAVLDIRGVSRATYSKDEFTLVK 120

DB 61 GDVKLTQSMATIRYIADKHNMLGSGPKRAEISMLEGAVLDIRGVSRATYSKDEFTLVK 120
OY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVVLYMDPMLDAFRLVSEK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVVLYMDPMLDAFRLVSEK 180
OY 181 KRIEAIPOIDKYLKSKSIYAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSKSIYAMPLOGWQATFGGDHPK 218

RESULT 14

US-08-216-894-4
Sequence 4, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRL0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-4

Query Match 98.7%; Score 1149; DB 2; Length 354;
Best Local Similarity 98.6%; Pred. No. 1.4e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLHYERDEGDKWNNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLHYERDEGDKWNNKKFELGLEFPNLPYYID 60
OY 61 GDVKLTQSMATIRYIADKHNMLGSGPKRAEISMLEGAVLDIRGVSRATYSKDEFTLVK 120
DB 61 GDVKLTQSMATIRYIADKHNMLGSGPKRAEISMLEGAVLDIRGVSRATYSKDEFTLVK 120
OY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVVLYMDPMLDAFRLVSEK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVVLYMDPMLDAFRLVSEK 180
OY 181 KRIEAIPOIDKYLKSKSIYAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSKSIYAMPLOGWQATFGGDHPK 218

```
RESULT 15
US-09-115-746-4
; Sequence 4, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; City: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DBLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-115-746-4

Query Match          98.7%; Score 1149; DB 4; Length 354;
Best Local Similarity 98.6%; Pred. No. 1,4e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYID 60
Db 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMATIRYIADKHNNLGGSPKRAEISMLEGAVLDIRYGVSRIRAYSKDFETLKV 120
Db 61 GDVKLTQSMATIRYIADKHNNLGGSPKRAEISMLEGAVLDIRYGVSRIRAYSKDFETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPKLYSPK 180
Db 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPKLYSPK 180
QY 181 KRIEAIPOIDKYLYKSSKIYAMPLOGWQATFGGGDHPK 218
Db 181 KRIEAIPOIDKYLYKSSKIYAMPLOGWQATFGGGDHPK 218
```

Search completed: May 20, 2003, 11:40:57
Job time : 8.02387 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 34.7335 Seconds

(Without alignments)
622,443 Million cell updates/sec

Title: US-10-081-408-5

Perfect score: 1164
Sequence: 1 MSPILGYWKIKGLVQPTRL.....IAMPLOQWQATFGGDDHPK 218

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PTI_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PTI_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1164	100.0	218	US-10-081-408-5	Sequence 5, Appli
2	1159	99.6	998	US-10-081-408-20	Sequence 20, Appli
3	1149	98.7	218	US-10-081-408-4	Sequence 4, Appli
4	1149	98.7	239	US-09-823-153-7	Sequence 7, Appli
5	1149	98.7	245	US-10-267-311-23	Sequence 23, Appli
6	1149	98.7	272	US-09-844-988-4	Sequence 4, Appli
7	1149	98.7	272	US-09-844-988-6	Sequence 6, Appli
8	1149	98.7	272	US-09-844-988-4	Sequence 4, Appli
9	1149	98.7	272	US-09-844-988-6	Sequence 6, Appli
10	1149	98.7	282	US-09-844-988-3	Sequence 3, Appli
11	1149	98.7	282	US-09-844-988-5	Sequence 5, Appli
12	1149	98.7	282	US-09-844-988-3	Sequence 3, Appli
13	1149	98.7	282	US-09-844-988-5	Sequence 5, Appli
14	1149	98.7	298	US-09-910-600-26	Sequence 26, Appli
15	1149	98.7	324	US-10-267-311-25	Sequence 25, Appli
16	1149	98.7	331	US-09-824-438-6	Sequence 6, Appli
17	1149	98.7	348	US-09-910-600-22	Sequence 22, Appli
18	1149	98.7	348	US-09-910-600-23	Sequence 23, Appli
19	1149	98.7	348	US-09-910-600-24	Sequence 24, Appli

20	1149	98.7	348	US-09-910-600-25	Sequence 25, Appli
21	1149	98.7	354	US-09-823-153-8	Sequence 8, Appli
22	1149	98.7	364	US-09-788-268-14	Sequence 14, Appli
23	1149	98.7	394	US-09-980-578-4	Sequence 4, Appli
24	1149	98.7	401	US-10-190-866A-1	Sequence 1, Appli
25	1149	98.7	412	US-09-775-964-34	Sequence 34, Appli
26	1149	98.7	422	US-09-824-438-7	Sequence 7, Appli
27	1149	98.7	1140	US-09-950-634-4	Sequence 4, Appli
28	1146	98.5	649	US-10-174-784-9	Sequence 9, Appli
29	483.5	41.5	229	US-10-102-806-554	Sequence 554, App
30	196.5	16.9	222	US-09-784-739-4	Sequence 4, Appli
31	196.5	16.9	230	US-09-925-301-1318	Sequence 1318, Ap
32	179.5	15.4	222	US-09-784-739-3	Sequence 3, Appli
33	169.5	14.6	203	US-09-847-208-48	Sequence 48, Appli
34	169.5	14.6	222	US-09-784-739-5	Sequence 5, Appli
35	161.5	13.9	222	US-09-784-739-1	Sequence 1, Appli
36	104	8.9	240	US-09-964-889-23	Sequence 23, Appli
37	103	8.8	233	US-09-765-213A-2	Sequence 2, Appli
38	96	8.2	233	US-09-765-213A-4	Sequence 4, Appli
39	95	8.2	233	US-09-765-213A-6	Sequence 6, Appli
40	90	7.7	379	US-10-253-007-46	Sequence 46, Appli
41	89	7.6	263	US-09-769-787-68	Sequence 68, Appli
42	88.5	7.4	359	US-09-738-626-4895	Sequence 4895, Ap
43	86	7.4	1668	US-09-815-242-5654	Sequence 5654, Ap
44	86	7.4	2397	US-09-815-242-12265	Sequence 12265, Ap
45	85.5	7.3	241	US-09-854-133-204	Sequence 204, App

ALIGNMENTS

RESULT 1
US-10-081-408-5
Sequence 5, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abirams, n, Lars
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: US 60/272,247
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 218
TYPE: PRT
ORGANISM: Schistosoma japonicum
US-10-081-408-5

Query Match 100.0%; Score 1164; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 3.4e-106;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSPILGYWKIKGLVQPTRLLEKYEHEHLYENDEGKRNKKFELGLEPNNLPYYD	60
DB	1	MSPILGYWKIKGLVQPTRLLEKYEHEHLYENDEGKRNKKFELGLEPNNLPYYD	60
QY	61	GVVKLTQSAAIRYADKNHMLGSPKREAEISMLEGAVLDIRYGSRIAYSKDEETLKV	120
DB	61	GVVKLTQSAAIRYADKNHMLGSPKREAEISMLEGAVLDIRYGSRIAYSKDEETLKV	120
QY	121	DFLSKLPKMLKMFEDRLSKTYLNGDHTVHPFMYDALDVLVYMDPKLAFPKLYSFK	180
DB	121	DFLSKLPKMLKMFEDRLSKTYLNGDHTVHPFMYDALDVLVYMDPKLAFPKLYSFK	180
QY	181	KRIEAIPOIDKYLKSSKYIAMPLOQWQATFGGDDHPK 218	
DB	181	KRIEAIPOIDKYLKSSKYIAMPLOQWQATFGGDDHPK 218	

RESULT 2

US-10-081-408-20
; Sequence 20, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrahams, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081,408
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 998
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant construct
US-10-081-408-20

Query Match 99.6%; Score 1159; DB 9; Length 998;
Best Local Similarity 100.0%; Pred. No. 7,7e-105;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SPILGYWKIKGLVOPTRLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
DB 38 SPILGYWKIKGLVOPTRLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
QY 62 DYKLTQSMALIRYIADKHNMLGSGPKERAISMEGAVLDIRYGVSRIVASKEDETLKY 121
DB 98 DYKLTQSMALIRYIADKHNMLGSGPKERAISMEGAVLDIRYGVSRIVASKEDETLKY 157
QY 122 FLSKLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLVYMPDMLDAFPKLVSEFK 181
DB 158 FLSKLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLVYMPDMLDAFPKLVSEFK 217
QY 182 RIEALPQIDKYLKSSKYIAMPLOGQATFGGDDHPPK 218
DB 218 RIEALPQIDKYLKSSKYIAMPLOGQATFGGDDHPPK 254

RESULT 3

US-10-081-408-4
; Sequence 4, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrahams, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081,408
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Schistosoma japonicum
US-10-081-408-4

Query Match 98.7%; Score 1149; DB 9; Length 218;
Best Local Similarity 98.6%; Pred. No. 9,9e-105;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTQSMALIRYIADKHNMLGSGPKERAISMEGAVLDIRYGVSRIVASKEDETLKY 120
DB 61 GGVKLTQSMALIRYIADKHNMLGSGPKERAISMEGAVLDIRYGVSRIVASKEDETLKY 120
QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLVYMPDMLDAFPKLVSEFK 180
DB 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLVYMPDMLDAFPKLVSEFK 180
QY 181 KRIEALPQIDKYLKSSKYIAMPLOGQATFGGDDHPPK 218
DB 181 KRIEALPQIDKYLKSSKYIAMPLOGQATFGGDDHPPK 218

RESULT 4

US-09-823-153-7
; Sequence 7, Application US/09823153
; Patent No. US20020025540A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; APPLICANT: Roberts, Susan
; APPLICANT: Lewis, Roger
; APPLICANT: Smith, David
; APPLICANT: Hendrick, Joseph
; APPLICANT: Vinitzky, Alexander
; TITLE OF INVENTION: ISOLATION OF FUNCTIONALLY ACTIVE GAMMA-SECRETASE PROTEIN COMPL
; FILE REFERENCE: D0004
; CURRENT APPLICATION NUMBER: US/09/823,153
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Description of Artificial Sequence: PSI PEPTIDE
US-09-823-153-7

Query Match 98.7%; Score 1149; DB 10; Length 239;
Best Local Similarity 98.6%; Pred. No. 1,1e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GGVKLTQSMALIRYIADKHNMLGSGPKERAISMEGAVLDIRYGVSRIVASKEDETLKY 120
DB 61 GGVKLTQSMALIRYIADKHNMLGSGPKERAISMEGAVLDIRYGVSRIVASKEDETLKY 120
QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLVYMPDMLDAFPKLVSEFK 180
DB 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLVYMPDMLDAFPKLVSEFK 180
QY 181 KRIEALPQIDKYLKSSKYIAMPLOGQATFGGDDHPPK 218
DB 181 KRIEALPQIDKYLKSSKYIAMPLOGQATFGGDDHPPK 218

RESULT 5

US-10-267-311-23
; Sequence 23, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:

REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-844-988-6

Query Match 98.7% Score 1149; DB 9; Length 272;
Best Local Similarity 98.6% Pred. No. 1.3e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYKIKGLVQPTLLLEYLEEKYEHLHYERDEGDKMKNKFFELGLEFPNLPYID 60
DB 1 MSPILGYKIKGLVQPTLLLEYLEEKYEHLHYERDEGDKMKNKFFELGLEFPNLPYID 60
QY 61 GDVKTQSMATIRYIADHNHMGSPKRAEISMLEGAVLDIRGVSRVIAYSKDEETLKV 120
DB 61 GDVKTQSMATIRYIADHNHMGSPKRAEISMLEGAVLDIRGVSRVIAYSKDEETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPPLVYFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPPLVYFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDHPPK 218

RESULT 8
US-09-844-908-4
Sequence 4, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gfan
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-908-4

Query Match 98.7% Score 1149; DB 10; Length 272;
Best Local Similarity 98.6% Pred. No. 1.3e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYKIKGLVQPTLLLEYLEEKYEHLHYERDEGDKMKNKFFELGLEFPNLPYID 60
DB 1 MSPILGYKIKGLVQPTLLLEYLEEKYEHLHYERDEGDKMKNKFFELGLEFPNLPYID 60
QY 61 GDVKTQSMATIRYIADHNHMGSPKRAEISMLEGAVLDIRGVSRVIAYSKDEETLKV 120
DB 61 GDVKTQSMATIRYIADHNHMGSPKRAEISMLEGAVLDIRGVSRVIAYSKDEETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPPLVYFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYMDPMLDAFPPLVYFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDHPPK 218

RESULT 9
US-09-844-908-6
Sequence 6, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gfan
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-844-988-6

Query Match 98.7%; Score 1149; DB 10; Length 272;

Best Local Similarity 98.6%; Pred. No. 1.3e-104;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
OY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKWKKFELGLEFPNLPYYID 60
    |||
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKWKKFELGLEFPNLPYYID 60
    |||
OY 61 GGVKLTQSAIIRIYADKNMGLGSPKERAETSMLEGAVALDIRGYSTRAYSKEDETLKY 120
    |||
DB 61 GGVKLTQSAIIRIYADKNMGLGSPKERAETSMLEGAVALDIRGYSTRAYSKEDETLKY 120
    |||
OY 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKCLDAFPKLVSPK 180
    |||
DB 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKCLDAFPKLVSPK 180
    |||
OY 181 KRLEAIPOIDKYLKSSKYIAMPLOGQWQATFGGDDHPK 218
    |||
DB 181 KRLEAIPOIDKYLKSSKYIAMPLOGQWQATFGGDDHPK 218
    |||
```

RESULT 10

US-09-844-988-3

Sequence 3, Application US/09844988

Patent No. US20020158764A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi

Barbosa, Miguel

Li, Gian

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESSES:

ADDRESS: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,988

FILING DATE: 26-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/910,820

FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-844-988-3

Query Match 98.7%; Score 1149; DB 9; Length 282;
Best Local Similarity 98.6%; Pred. No. 1.4e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
OY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKWKKFELGLEFPNLPYYID 60
    |||
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKWKKFELGLEFPNLPYYID 60
    |||
OY 61 GGVKLTQSAIIRIYADKNMGLGSPKERAETSMLEGAVALDIRGYSTRAYSKEDETLKY 120
    |||
DB 61 GGVKLTQSAIIRIYADKNMGLGSPKERAETSMLEGAVALDIRGYSTRAYSKEDETLKY 120
    |||
OY 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKCLDAFPKLVSPK 180
    |||
DB 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKCLDAFPKLVSPK 180
    |||
OY 181 KRLEAIPOIDKYLKSSKYIAMPLOGQWQATFGGDDHPK 218
    |||
DB 181 KRLEAIPOIDKYLKSSKYIAMPLOGQWQATFGGDDHPK 218
    |||
```

RESULT 11

US-09-844-988-5

Sequence 5, Application US/09844988

Patent No. US20020158764A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi

Barbosa, Miguel

Li, Gian

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESSES:

ADDRESS: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,988

FILING DATE: 26-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/910,820

FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-844-988-5

Query Match 98.7%; Score 1149; DB 9; Length 282;
Best Local Similarity 98.6%; Pred. No. 1.4e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVQPTLLEYLEEKEEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTLLEYLEEKEEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
QY 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAVSKDEFTLVK 120
DB 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAVSKDEFTLVK 120
QY 121 DFLSLPEMLKFEEDRLSHKTYLNGDHYTHPDMFLYDALDVLYLMDPKCLDAFPKLVYFK 180
DB 121 DFLSLPEMLKFEEDRLSHKTYLNGDHYTHPDMFLYDALDVLYLMDPKCLDAFPKLVYFK 180
QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGMQATFFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGMQATFFGGGDHPK 218

RESULT 12
US-09-844-908-3
Sequence 3, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gfan
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-844-908-3

Query Match 98.7%, Score 1149, DB 10, Length 282;
Best Local Similarity 98.6%, Pred. No. 1,4e-104;
Matches 215, Conservative 0, Mismatches 3, Indels 0, Gaps 0;
QY 1 MSPILGWKIKGLVQPTLLEYLEEKEEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTLLEYLEEKEEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60

QY 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAVSKDEFTLVK 120
DB 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAVSKDEFTLVK 120
QY 121 DFLSLPEMLKFEEDRLSHKTYLNGDHYTHPDMFLYDALDVLYLMDPKCLDAFPKLVYFK 180
DB 121 DFLSLPEMLKFEEDRLSHKTYLNGDHYTHPDMFLYDALDVLYLMDPKCLDAFPKLVYFK 180
QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGMQATFFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGMQATFFGGGDHPK 218

RESULT 13
US-09-844-908-5
Sequence 5, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gfan
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-844-908-5

Query Match 98.7%, Score 1149, DB 10, Length 282;
Best Local Similarity 98.6%, Pred. No. 1,4e-104;
Matches 215, Conservative 0, Mismatches 3, Indels 0, Gaps 0;
QY 1 MSPILGWKIKGLVQPTLLEYLEEKEEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTLLEYLEEKEEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
QY 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAVSKDEFTLVK 180
DB 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAVSKDEFTLVK 180

QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHPDMFLYDALDVLVYMDPMLCLDAFPKLVSEK 180
 |||||
 Db 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLVYMDPMLCLDAFPKLVSEK 180
 |||||
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218
 |||||
 Db 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218

RESULT 14

US-09-910-600-26
 ; Sequence 26, Application US/09910600
 ; Publication No. US2003003631A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Longphre, Malinda
 ; APPLICANT: Chang, Han
 ; APPLICANT: Whitley, Gena
 ; TITLE OF INVENTION: NOVEL SINGLETS AND USES THEREOF
 ; FILE REFERENCE: D0003NP
 ; CURRENT APPLICATION NUMBER: US/09/910,600
 ; CURRENT FILING DATE: 2001-07-20
 ; PRIOR APPLICATION NUMBER: 60/220,139
 ; PRIOR FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 26
 ; LENGTH: 298
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: L3cyto-Y641
 ;
 US-09-910-600-26

Query Match 98.7%; Score 1149; DB 9; Length 298;
 Best Local Similarity 98.6%; Pred. No. 1.5e-104;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVOPTRLLLEYLEEKYEHLHYERDEGDKWRNKKFELGLEFPNLPYYID 60
 |||||
 Db 1 MSPILGWKIKGLVOPTRLLLEYLEEKYEHLHYERDEGDKWRNKKFELGLEFPNLPYYID 60
 |||||
 QY 61 GDVKLTOSMAIIRYIADKHNMLGGSPKRAEISMLEGAVLDIRYGVSRAYSKDPEITLV 120
 |||||
 Db 61 GDVKLTOSMAIIRYIADKHNMLGGSPKRAEISMLEGAVLDIRYGVSRAYSKDPEITLV 120
 |||||
 QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHPDMFLYDALDVLVYMDPMLCLDAFPKLVSEK 180
 |||||
 Db 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLVYMDPMLCLDAFPKLVSEK 180
 |||||
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218
 |||||
 Db 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218

RESULT 15

US-10-267-311-25
 ; Sequence 25, Application US/10267311
 ; Publication No. US20030050469A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Siegel, Marvin
 ; APPLICANT: Chu, N. Randall
 ; APPLICANT: Mizen, Lee A.
 ; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
 ; FILE REFERENCE: 12071/002001
 ; CURRENT APPLICATION NUMBER: US/10/267,311
 ; CURRENT FILING DATE: 2002-10-09
 ; PRIOR APPLICATION NUMBER: US/09/613,303
 ; PRIOR FILING DATE: 2000-07-10
 ; PRIOR APPLICATION NUMBER: US 60/143,757
 ; PRIOR FILING DATE: 1999-07-08
 ; NUMBER OF SEQ ID NOS: 55

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 25
 ; LENGTH: 324
 ; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:
 ; OTHER INFORMATION: fusion sequence

US-10-267-311-25

Query Match 98.7%; Score 1149; DB 9; Length 324;
 Best Local Similarity 98.6%; Pred. No. 1.7e-104;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVOPTRLLLEYLEEKYEHLHYERDEGDKWRNKKFELGLEFPNLPYYID 60
 |||||
 Db 1 MSPILGWKIKGLVOPTRLLLEYLEEKYEHLHYERDEGDKWRNKKFELGLEFPNLPYYID 60
 |||||
 QY 61 GDVKLTOSMAIIRYIADKHNMLGGSPKRAEISMLEGAVLDIRYGVSRAYSKDPEITLV 120
 |||||
 Db 61 GDVKLTOSMAIIRYIADKHNMLGGSPKRAEISMLEGAVLDIRYGVSRAYSKDPEITLV 120
 |||||
 QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHPDMFLYDALDVLVYMDPMLCLDAFPKLVSEK 180
 |||||
 Db 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLVYMDPMLCLDAFPKLVSEK 180
 |||||
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218
 |||||
 Db 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218

Search completed: May 20, 2003, 12:02:22

Job time : 35.7335 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 5.14571 Seconds

(Without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-5

Perfect score: 1164

Sequence: 1 MSPILGYWKIKGLVQPTRL.....IAMPLOGWQATFGGDPK 218

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwisProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	98.7	218	1	GT26_SCHJA
2	990	85.1	218	1	GT27_SCHJA
3	927	79.6	218	1	GT26_SCHMA
4	717.5	61.6	220	1	GT29_FASHE
5	708	60.8	217	1	GT27_FASHE
6	690	59.3	217	1	GT28_FASHE
7	689	59.2	217	1	GT26_FASHE
8	489.5	42.1	217	1	GTM1_MOUSE
9	489.5	42.1	217	1	GTM1_RAT
10	488.5	42.0	218	1	GTM4_HUMAN
11	486.5	41.8	218	1	GTM6_MOUSE
12	485.5	41.7	217	1	GTM2_MOUSE
13	484.5	41.6	217	1	GTM2_HUMAN
14	481.5	41.4	217	1	GTM2_MOUSE
15	476.5	40.9	217	1	GTM2_MOUSE
16	476.5	40.9	217	1	GTM2_MOUSE
17	475.5	40.9	217	1	GTM2_MOUSE
18	473.5	40.7	217	1	GTM2_MOUSE
19	471.5	40.5	217	1	GTM2_MOUSE
20	469.5	40.3	217	1	GTM2_MOUSE
21	463.5	39.8	217	1	GTM2_MOUSE
22	460.5	39.6	217	1	GTM2_MOUSE
23	454.5	39.0	224	1	GTM3_HUMAN
24	443.5	38.1	224	1	GTM3_MOUSE
25	440.5	37.8	219	1	GTM2_CHICK
26	394.5	33.9	219	1	GTM1_MOUSE
27	259.5	22.3	208	1	GTM1_MOUSE
28	254.5	21.9	208	1	GTM1_MOUSE
29	253	21.9	208	1	GTM1_MOUSE
30	252	21.6	207	1	GTM1_MOUSE
31	251	21.6	207	1	GTM1_MOUSE
32	249	21.4	209	1	GTM1_MOUSE
33	247	21.2	209	1	GTM1_MOUSE

34	247	21.2	209	1	GTP_CRIMI
35	246	21.1	209	1	GTP_RAT
36	245	21.0	209	1	GTP_CRILLO
37	244	21.0	209	1	GTP_MACMU
38	243	20.9	209	1	GTP_MOUSE
39	239	20.5	209	1	GTP_BOVIN
40	226.5	19.5	208	1	GTP_CAERL
41	221.5	19.0	223	1	GTAL_RABIT
42	213.5	18.3	210	1	GTP2_BUFB
43	209.5	18.0	221	1	GTAL_RAT
44	207.5	17.8	221	1	GTAL_MOUSE
45	207.5	17.8	222	1	GTAL_MOUSE

ALIGNMENTS

RESULT 1	ID	GT26_SCHJA	STANDARD;	PRT;	218 AA.
AC	P08515;				
DT	01-AUG-1988	(Rel. 08, Created)			
DT	01-AUG-1988	(Rel. 39, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (S26 antigen)				
DE	(GST class-alpha).				
OS	Schistosoma japonicum (Blood fluke).				
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoida;				
OC	Schistosomatidae; Schistosomatidae; Schistosoma.				
OX	NCBI_Taxid=6182;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=87041520; PubMed=3095841;				
RA	Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,				
RA	Mitchell G.F.;				
RA	Proc. Natl. Acad. Sci. U.S.A. 84:6541-6541(1987).				
RN	[3]				
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).				
RX	MEDLINE=9527631; PubMed=7538846;				
RA	Liu K., Ho J.X., Keeling K., Gilliland G.L., Ji X., Ruker F.,				
RA	Carter D.C.;				
RT	"Three-dimensional structure of Schistosoma japonicum glutathione S-				
RT	transferase fused with a six-amino acid conserved neutralizing				
RT	epitope of gp41 from HIV."				
RL	Protein Sci. 3:2233-2244(1994).				
RN	[4]				
RP	X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).				
RX	MEDLINE=9515648; PubMed=7853399;				
RA	McGlynn M.A., Williams D.R., Tainer J.A.;				
RT	"Crystal structures of a schistosomal drug and vaccine target:				
RT	glutathione S-transferase from Schistosoma japonica and its complex				
RT	with the leading antischistosomal drug praziquantel."				
RL	J. Mol. Biol. 246:21-27(1995).				
CC	-1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER				
CC	OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.				
CC	-1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE				
CC	PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO				
CC	SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF				
CC	HAEMATIN IN THE PARASITE GUT.				
CC	-1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.				
CC	-1- SUBUNIT: HOMODIMER.				
CC	-1- MISCELLANEOUS: THERE ARE AT LEAST TWO ISOENZYMES OF GST IN				
CC	S. JAPONICUM.				
CC	-1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL/ M14654; AAB59203.1; -
DR PIR; A26484; A26484.
DR PDB; ICGNE; 30-NOV-94.
DR PDB; IGTB; 07-FEB-95.
DR PDB; IGTB; 01-DEC-95.
DR PDB; 18BX; 12-APR-99.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR Transferrase; Antigen; Multigene family; 3D-structure.
KW SEQUENCE 218 AA; 25498 MW; 5E2AC418BDEE13F CRC64;

Query Match 98.7%; Score 1149; DB 1; Length 218;
Best Local Similarity 98.6%; Pred. No. 3.8e-90;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVQPTLLLEYEEKYEERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTLLLEYEEKYEERDEGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GDVRLTQSMALIRYIADKHNMLGSSPKERAISMLEGAVDIRGVSRIVASKDFETLKY 120
DB 61 GDVRLTQSMALIRYIADKHNMLGSSPKERAISMLEGAVDIRGVSRIVASKDFETLKY 120
QY 121 DFLSKLPBMLKMFEDRLSHKTYTLNGDHYTHDPFMYLDALDVLVYMDPCIDAFKLYSFK 180
DB 121 DFLSKLPBMLKMFEDRLSHKTYTLNGDHYTHDPFMYLDALDVLVYMDPCIDAFKLYSFK 180
QY 121 DFLSKLPBMLKMFEDRLSHKTYTLNGDHYTHDPFMYLDALDVLVYMDPCIDAFKLYSFK 180
DB 121 DFLSKLPBMLKMFEDRLSHKTYTLNGDHYTHDPFMYLDALDVLVYMDPCIDAFKLYSFK 180
QY 181 KRFAIPQIDKYLKSKSYIAPLOGMOATFGGDPK 218
DB 181 KRFAIPQIDKYLKSKSYIAPLOGMOATFGGDPK 218

RESULT 2
GT27_SCHMA STANDARD; PRT; 218 AA.
ID GT27_SCHMA STANDARD; PRT; 218 AA.
AC P35661;
DT 01-JUN-1994 (Rel. 29) Created
DT 01-JUN-1994 (Rel. 29) Last sequence update
DT 01-NOV-1995 (Rel. 32, last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/2
DE antigen) (GST class-alpha).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92131046; PubMed=1775156;
RA Wright M.D., Harrison R.A., Melder A.M., Newport G.R., Mitchell G.F.;
RT "Another 26-kilodalton glutathione S-transferase of Schistosoma
RT mansoni".
RL Mol. Biochem. Parasitol. 49:177-179(1991).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL/ M73624; -; NOT_ANNOTATED_CDs.
DR PIR; A45556; A45556.
DR HSSP; P08515; IGTa.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR Transferrase; Antigen; Multigene family.
KW SEQUENCE 218 AA; 25411 MW; D8D3EE9028B36185 CRC64;

Query Match 85.1%; Score 990; DB 1; Length 218;
Best Local Similarity 83.0%; Pred. No. 1e-76;
Matches 181; Conservative 19; Mismatches 18; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVQPTLLLEYEEKYEERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MAPRLGYWKIKGLVQPTLLLEYEEKYEERDEGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GDVRLTQSMALIRYIADKHNMLGSSPKERAISMLEGAVDIRGVSRIVASKDFETLKY 120
DB 61 GDVRLTQSMALIRYIADKHNMLGSSPKERAISMLEGAVDIRGVSRIVASKDFETLKY 120
QY 121 DFLSKLPBMLKMFEDRLSHKTYTLNGDHYTHDPFMYLDALDVLVYMDPCIDAFKLYSFK 180
DB 121 DFLSKLPBMLKMFEDRLSHKTYTLNGDHYTHDPFMYLDALDVLVYMDPCIDAFKLYSFK 180
QY 121 DFLSKLPBMLKMFEDRLSHKTYTLNGDHYTHDPFMYLDALDVLVYMDPCIDAFKLYSFK 180
DB 121 DFLSKLPBMLKMFEDRLSHKTYTLNGDHYTHDPFMYLDALDVLVYMDPCIDAFKLYSFK 180
QY 181 KRFAIPQIDKYLKSKSYIAPLOGMOATFGGDPK 218
DB 181 KRFAIPQIDKYLKSKSYIAPLOGMOATFGGDPK 218

RESULT 3
GT26_SCHMA STANDARD; PRT; 218 AA.
ID GT26_SCHMA STANDARD; PRT; 218 AA.
AC P15964;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/1
DE antigen) (GST class-alpha).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Puerto Rican;
RX MEDLINE=90348716; PubMed=2385266;
RA Trolein F., Kieny M.P., Vervaeke C., Torgler G., Pierce R.J.,
RA Balloul J.-M., Schmitt D., Lecocq J.-P., Capron A.;
RT "Molecular cloning and tissue distribution of a 26-kilodalton
RT Schistosoma mansoni glutathione S-transferase".
RL Mol. Biochem. Parasitol. 41:35-44(1990).
RN [2]
RP SEQUENCE OF 8-218 FROM N.A.
RX STRAIN-Puerto Rican;
RX MEDLINE=90271935; PubMed=1693415;
RA Herkle K.J., Davern K.M., Wright M.D., Ramos A.J., Mitchell G.F.;
RT "Comparison of the cloned genes of the 26- and 28-kilodalton
RT glutathione S-transferases of Schistosoma japonicum and Schistosoma
RT mansoni".
RL Mol. Biochem. Parasitol. 40:23-34(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF

```

RN Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 22-220 FROM N.A.
RA Cramerit S.;
RL Patent number WO9008819, 09-AUG-1990.
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: A00993; CAA00118.1; -.
DR HSSP: P31670; IPIE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C: 1.
DR Pfam: PF02798; GST_N: 1.
KW Transferase; Antigen; Multigene family.
FT INIT_MET 0 BY SIMILARITY.
FT CONFLICT 22 22 Y->V (IN REF. 3).
FT CONFLICT 110 111 DP->VS (IN REF. 3).
FT CONFLICT 189 189 A->P (IN REF. 3).
SQ SEQUENCE 220 AA; 25598 MW; 27B9F150B75D101F CRC64;

Query Match 61.6%; Score 717.5; DB 1; Length 220;
Best Local Similarity 60.6%; Pred. No. 1e-53;
Matches 129; Conservative 32; Mismatches 51; Indels 1; Gaps 1

OY 5 LGYWKIKGLVQPTRLLEYLEEKYDEHLYREBQDGKWRMKRELGLEPNLPYYIDGVK 64
DB 4 LGYWKIKGLQGVRLLEL-GEKYEQIYERDQGEKWSKKEGLDLPNLPYYIDDKK 62
OY 65 LTQSAIIRYIADKNHMLGSPKREAEISMLEGAVLDIRGVYSRIAYSKPFTLKYVPLS 124
DB 63 LTQSLAIRYIADKNIGMIGSPREPARARVSMIEGAVALDNLQGLSRISYDPKFLQEGYLK 122
OY 125 KLPKLMKMFEDRLSKRTYINGDHVTHPPDMLYDALDLYLNDPNCIDAPFKLYSFKRRI 184
DB 123 DLPTTKMMSDPLGKNRPYLGTSVSHVDPMYEALDAIRYLEPHCDHPNPLOQFMSRI 182
OY 185 AIPQIDKYLKSSKYTAMPLOQMGQAFHGGGDHP 217
DB 183 ALPSIKATWESNRFITKWPPLNGWHAQGGGDAPP 215

RESULT 5
GT27_FASHE
ID GT27_FASHE STANDARD; PRT; 217 AA.
AC P31670.
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GSTf7) (F447)
DE (GST class-alpha).
DE Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenae;
OC Echinostomida; Echinostomata; Fascioloidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;

```

RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of *Fasciola hepatica*.";
 RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RP SEQUENCE OF 7-105 FROM N.A.
 RA Crameri S.;
 RL Patent number WO9008819, 09-AUG-1990.
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
 RX MEDLINE=98035725; PubMed=9367777;
 RA Rosjohn J., Fell S.C., Milce M.C.J., Sexton J.L., Spithill T.W.,
 RA Parker M.W.;
 RT "Crystallization, structural determination and analysis of a novel
 RT parasite vaccine candidate: *Fasciola hepatica* glutathione
 RT S-transferase.";
 RL J. Mol. Biol. 273:857-872(1997).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasm;c.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M7681; AAA29140.1; -;
 DR EMBL; A00996; CA00121.1; -;
 DR PDB; 1FHE; 29-JUL-98.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C_1.
 DR Pfam: PF02798; GST_N_1.
 DR Transferrase; Antigen: Multigene family; 3D-structure.
 KW INT_MET 0
 FT CONFLICT 65 65 T -> I (IN REF. 3).
 FT CONFLICT 102 105 RIGF -> FEEL (IN REF. 3).
 FT SEQUENCE 217 AA; 25281 MW; 0FB8BFE63029E03 CRC64;
 SQ
 Query Match 60.8%; Score 708; DB 1; Length 217;
 Best Local Similarity 58.2%; Pred. No. 6.5e-53;
 Matches 124; Conservative 37; Mismatches 52; Indels 0; Gaps 0;
 OY 5 LGYWKIKGLVPTRLLEVEEYEHLYERDEGKMRKKKELEFPNLYYIDGVK 64
 DB 4 LGYWKIRGLQPVRLLEVEEYEHLYGRDRKMKSEKMGIDLPNLYYIDDKK 63
 OY 65 LTQSAIIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRAYSKEFETLVDPDS 124
 DB 64 LTQSAIIRYIADKHNMLGSGTPERARISMTIGAMDLRIGGRVCYNKFEVVAEEYK 123
 OY 125 KLPEMLKMFEDRLSHKTYLNGDVTHTPDFMLYDALDVLVYMDPCLDAEPKLVSEKRIE 184
 DB 124 ELPKTLKMSDFLDGRHYLTGSSVSHVDFMLYETLDSIRYLAHPCIDEPKLVSEKRIE 183
 OY 185 AIPQIDKYLKSKYIAMPPLQGWATFFGGDHP 217
 DB 184 ALPKIKAYMESKRLTKWPLNGMAASFAGADAP 216

RESULT 6
 GT28_FASHE STANDARD: PRT; 217 AA.
 ID ID FASHE
 AC P31671;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 7 (EC 2.5.1.18) (GST7) (FH7) (GST
 DE class-alpha).
 OS *Fasciola hepatica* (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fascioloidae; Fasciolidae; Fasciola.
 OX NCBI_Taxid=6192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92155306; PubMed=1740183;
 RX Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of *Fasciola hepatica*.";
 RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RP SEQUENCE OF 8-217 FROM N.A.
 RA Crameri S.;
 RL Patent number WO9008819, 09-AUG-1990.
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M7680; AAA29139.1; -;
 DR EMBL; A00994; CA00119.1; -;
 DR HSRP; P31670; 1FHE.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C_1.
 DR Pfam: PF02798; GST_N_1.
 DR Transferrase; Antigen: Multigene family.
 KW INT_MET 0
 FT CONFLICT 35 42 NDREKWLQ -> MGRNGMA (IN REF. 3).
 FT CONFLICT 188 196 IKYWKSR -> SYKMSRA (IN REF. 3).
 FT SEQUENCE 217 AA; 25196 MW; 0099ELF59E49A49E CRC64;
 SQ
 Query Match 59.3%; Score 690; DB 1; Length 217;
 Best Local Similarity 58.7%; Pred. No. 2.1e-51;
 Matches 125; Conservative 30; Mismatches 58; Indels 0; Gaps 0;
 OY 5 LGYWKIKGLVPTRLLEVEEYEHLYERDEGKMRKKKELEFPNLYYIDGVK 64
 DB 4 LGYWKIRGLQPVRLLEVEEYEHLYGRDRKMKSEKMGIDLPNLYYIDDKK 63
 OY 65 LTQSAIIRYIADKHNMLGSGPKERAETISMLEGAVLDIRYGVSRAYSKEFETLVDPDS 124
 DB 64 LTQSAIIRYIADKHNMLGSGTPERARISMTIGAMDLRIGGLTCYNPKFEELKGYLK 123

RA Mannervik B., Alln P., Gutenberg C., Jansson H., Tahlr M.K.,
 RA Marholm M., Joernvall H.;
 RT "Identification of three classes of cytosolic glutathione transferase
 RT common to several mammalian species: correlation between structural
 RT data and enzymatic properties.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7202-7206(1985).
 RN [7]
 RP CHARACTERIZATION.
 RP STRAIN-CD-1; TISSUE=Liver;
 RX MEDLINE-96189427; Pubmed-8605288;
 RA Mitchell A.E., Morin D., Lame M.W., Jones A.D.;
 RT "Purification, mass spectrometric characterization, and covalent
 RT modification of murine glutathione S-transferases";
 RL Chem. Res. Toxicol. 8:1054-1062(1995).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MASS SPECTROMETRY: MW-25838.4; MW-ERR-2; METHOD=Electrospray.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J03952; AAA37747.1; -;
 DR EMBL: J04632; AAA37705.1; -;
 DR EMBL: L13448; -; NOT_ANNOTATED_CDS.
 DR EMBL: BC003822; AAH03822.1; -;
 DR PIR: A20831; A20831.
 DR PIR: A28946; A28946.
 DR PIR: I24735; I24735.
 DR PIR: A34159; A34159.
 DR PIR: S33860; S33860.
 DR HSSP: P04905; 2GST.
 DR SWISS-2DPAGE: P10649; MOUSE.
 DR MGD: MGI:95860; Gstm1.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST_C1.
 DR Pfam: PF02798; GST_N1.
 DR PRINTS: PRO1267; GSTRSPASEM.
 DR Transferase: Multigene family.
 KM INIT_MET 0
 FT INIT_MET 0
 SQ SEQUENCE 217 AA; 25839 MW; A1EE3938F590B829 CRC64;
 Query Match 42.18; Score 489.5; DB 1; Length 217;
 Best Local Similarity 44.48; Pred. No. 1.8e-34;
 Matches 91; Conservative 39; Mismatches 70; Indels 5; Gaps 1;
 QY 4 ILGYWKIGVOPRLLEYLEEYKEHLYERDEG-----DKMNKKFELGLEFPNPLY 58
 DB 3 ILGVWNRGLHPIRMLEETDSDYDEKRTMGADPDGDRSQMLNEKFKLCLDPPNLYL 62
 QY 59 IDGVKLTQSMALIRYADKHNMLGGSPKERAELISMLEGAVLDIRYGVSRISAKDEFTL 118
 DB 63 IDGSHKKTQSNALIRYADKHNMLGGSPKERAELISMLEGAVLDIRYGVSRISAKDEFTL 122
 QY 119 KVDPLSLPEMLKMEEDRLSHKTYLNDGHTHPFMYLDALDVLVMDPKLDFPKLYS 178
 DB 123 KPEELKTIPEMKLYSFLKRPFPADKVTYVDLAVYDIDYRMEPEPKCLDAFPLNRD 182
 QY 179 FKRIEAIPOIDKYLKSSKYIAMPL 203
 DB 183 FLARFELKKISAMKSSRIATPI 207

RESULT 9
 GTM1_RAT
 ID GTM1_RAT STANDARD; PRT; 217 AA.
 AC P04905;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glutathione S-transferase Yb1 (EC 2.5.1.18) (Chain 3) (GST M1-1)
 DE (GST class-mu 1).
 GN GSTM1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (CLONE PGTR200).
 RX MEDLINE-86312882; Pubmed-2875437;
 RA Lai H.-C.J., Grove G., Tu C.-P.D.;
 RT "Cloning and sequence analysis of a cDNA for a rat liver glutathione
 RT S-transferase Yb subunit";
 RL Nucleic Acids Res. 14:6101-6114(1986).
 RN [2]
 RP SEQUENCE FROM N.A. (CLONE PGTA/C44).
 RX MEDLINE-86033768; Pubmed-3840477;
 RA Ding G.-J.-F., Lu A.Y.H., Pickett C.B.;
 RT "Rat liver glutathione S-transferases. Nucleotide sequence analysis
 RT of a Yb1 cDNA clone and prediction of the complete amino acid
 RT sequence of the Yb1 subunit";
 RL J. Biol. Chem. 260:13268-13271(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-86224097; Pubmed-3011803;
 RA Ding G.-J.-F., Ding V.D.-H., Rodkey J.A., Bennett C.D., Lu A.Y.H.,
 RA Pickett C.B.;
 RT "Rat liver glutathione S-transferases. DNA sequence analysis of a Yb2
 RT cDNA clone and regulation of the Yb1 and Yb2 mRNAs by
 RT phenobarbital";
 RL J. Biol. Chem. 261:7952-7957(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87308179; Pubmed-3040722;
 RA Chang C., Saltzman A.G., Sorensen N.S., Hlipakka R.A., Liao S.;
 RT "Identification of glutathione S-transferase Yb1 mRNA as the
 RT androgen-repressed mRNA by cDNA cloning and sequence analysis.";
 RL J. Biol. Chem. 262:11901-11903(1987).
 RN [5]
 RP SEQUENCE OF 1-23.
 RP STRAIN-Wistar; TISSUE=Olfactory epithelium;
 RX MEDLINE-93277499; Pubmed-8503873;
 RA Ben-Arie N., Khen M., Lancet D.;
 RT "Glutathione S-transferases in rat olfactory epithelium:
 RT purification, molecular properties and odorant biotransformation.";
 RL Biochem. J. 292:379-384(1993).
 RN [6]
 RP MUTAGENESIS OF CYS-86.
 RX MEDLINE-91354218; Pubmed-1883338;
 RA Hsieh J.-C., Huang S.-C., Chen W.-L., Lai Y.-C., Tam M.F.;
 RT "Cysteine 86 is not needed for the enzymic activity of glutathione S-
 RT transferase 3-3.";
 RL Biochem. J. 278:293-297(1991).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE-93041702; Pubmed-1420139;
 RA Ji X., Zhang P., Armstrong R.N., Gilliland G.L.;
 RT "The three-dimensional structure of a glutathione S-transferase from
 RT the yme gene class. Structural analysis of a glutathione S-transferase
 RT isoenzyme 3-3 and glutathione at 2.2-A resolution";
 RL Biochemistry 31:10169-10184(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RA Fu J.-H., Rose J., Tam M.F., Wang B.-C.;
 RT "New crystal forms of a mu-class glutathione S-transferase from rat
 RT liver.";

RL Acta Crystallogr. D 50:219-224(1994).
 RN (9)
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE-94153886; PubMed-8110735;
 RA Ji X., Johnson W.W., Sesay M.A., Dickert L., Prasad S.M., Ammon H.L.,
 RA Armstrong R.N., Gilliland G.L.;
 RT "Structure and function of the xenobiotic substrate binding site of a
 RT glutathione S-transferase as revealed by X-ray crystallographic
 RT analysis of product complexes with the diastereomers of 9-(S-
 RT glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene";
 RT Biochemistry 33:1043-1052(1994).
 CC -1- FUNCTION: CONFIGURATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: THE OLFACTORY GST MAY BE CRUCIAL FOR THE ACTIVITY OF THE
 CC OLFACTORY PROCESS.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MISCELLANEOUS: RAT LIVER GLUTATHIONE S-TRANSFERASES (GST) CONSIST
 CC OF BINARY COMBINATIONS OF THREE MAJOR CLASSES OF SUBUNITS
 CC DESIGNATED: YA (25 KDA), YB (27 KDA) AND YC (28 KDA).
 CC -1- MISCELLANEOUS: YB SUBCLASS SELECTIVELY BINDS STEROID HORMONES.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X04229; CA27811.1;
 DR EMBL: M1719; AAA41287.1;
 DR EMBL: J02810; AAA41293.1;
 DR PIR: A24085; A24085.
 DR PIR: A25510; A25510.
 DR PIR: A29794; A29794.
 DR PIR: S17167; S17167.
 DR PDB: 1G5B; 31-OCT-93.
 DR PDB: 1G5C; 31-OCT-93.
 DR PDB: 2G5T; 31-OCT-93.
 DR PDB: 3G5T; 31-OCT-93.
 DR PDB: 4G5T; 31-OCT-93.
 DR PDB: 5G5T; 31-OCT-93.
 DR PDB: 6G5T; 08-NOV-96.
 DR PDB: 6G5U; 08-NOV-96.
 DR PDB: 6G5V; 08-NOV-96.
 DR PDB: 6G5W; 08-NOV-96.
 DR PDB: 6G5X; 08-NOV-96.
 DR PDB: 6G5Y; 08-NOV-96.
 DR PDB: 5FMG; 27-JAN-99.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Pfam: PF02798; GST_mu.
 DR Transfaserse: Multigene family; 3D-structure; Olfaction.
 KM TRANSFASERSE: Multigene family; 3D-structure; Olfaction.
 FT INIT MET 0 0
 FT MUTAGEN 86 86 C->S: NO CHANGE IN ACTIVITY.
 FT CONFLICT 168 168 I->N (IN REF. 3).
 FT CONFLICT 198 199 KS->NC (IN REF. 2).
 FT STRAND 2 7
 FT TURN 11 13
 FT HELIX 14 22
 FT TURN 23 24
 FT STRAND 27 32
 FT TURN 37 39
 FT HELIX 43 46
 FT TURN 47 50
 FT STRAND 61 64
 FT TURN 65 66

FT STRAND 67 70
 FT HELIX 72 82
 FT TURN 83 84
 FT HELIX 90 114
 FT TURN 115 115
 FT TURN 117 118
 FT HELIX 119 128
 FT TURN 129 129
 FT HELIX 130 141
 FT TURN 142 143
 FT TURN 154 154
 FT HELIX 155 169
 FT TURN 171 176
 FT HELIX 178 188
 FT TURN 189 189
 FT HELIX 191 196
 FT TURN 197 198
 FT TURN 200 201
 FT TURN 210 211
 FT TURN 217 217
 SQ SEQUENCE 217 AA; 25782 MW; 2AC8BD9DA785118 CRC64;
 Query Match 42.1%; Score 489.5; DB 1; Length 217;
 Best Local Similarity 44.4%; Pred. No. 1.8e-34;
 Matches 91; Conservative 39; Mismatches 70; Indels 5; Gaps 1;
 QY 4 ILGYWKIKGLVOPTRLLLEYLEEYEHLYERDEG-----DKRNKKFELGLEPNIPLY 58
 DB 3 ILGYWNNRYGTHPRILLETDTSSYEKKRYAMGADPDYDRSSQWNEFKGLDPLNPLYL 62
 QY 59 IDGDKVLTQSMATIRYADKHNMLGSPKERAETSMGAVLDIRYGVSRIVASKDFETL 118
 DB 63 IDGSRKRTQSNALMKRYLARKHNLGCTEERIRADYENOVNRMQLINICNPDEKQ 122
 QY 119 KVDFLSKLPMLKMFEDRLSHKTYLNGDHYTHPDMLYDALDVLVYMDPMLDAFPLVS 178
 DB 123 KPEFLKTIPEKMKLYSEFLKRPWAGDKVYVDFLAYDILDQYHIEPRCLDAFPMLKD 182
 QY 179 FKRIEAIPIQIDKYLKSSXYIAMP 203
 DB 183 FLARFEGIKRISAYMKSSRYLSTPI 207
 RESULT 10
 ID GTM4_HUMAN STANDARD; PRT; 218 AA.
 AC Q03013;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glutathione S-transferase Mu 4 (EC 2.5.1.18) (GSTM4-4) (GTS-Mu2)
 DE (GST class-mu 4).
 GN GSTM4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid:9606;
 OX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93352467; PubMed-8349586;
 RA Comstock K.E., Johnson K.J., Rifkenbery D., Hennen W.D.;
 RT "Isolation and analysis of the gene and cDNA for a human Mu class
 RT glutathione S-transferase, GSTM4";
 RL J. Biol. Chem. 268:16958-16965(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93228631; PubMed-8471052;
 RA Zhong S., Spurr N.K., Hayes J.D., Wolf C.R.;
 RT "Deduced amino acid sequence, gene structure and chromosomal location
 RT of a novel human class Mu glutathione S-transferase, GSTM4";
 RL Biochem. J. 291:41-50(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Lung;


```

RP SEQUENCE.
RC TISSUE=Liver.
RX MEDLINE-90235961; PubMed-2332413;
RA Kamei K., Oshino R., Hara S.;
RT "Amino acid sequence of glutathione S-transferase b from guinea pig
  liver."
RL J. Biochem. 107:111-117(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
  OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- CATALYTIC ACTIVITY: Rx + glutathione - Hx + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
DR PIR: JX0095.
DR HSSP: P04905. 2GST.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR InterPro: IPR003081; GST_mu.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR PRINTS: PR01267; GSTRNSFRASEM.
KW Transferase; Multigene family.
SQ SEQUENCE 217 AA; 25719 MW; D29F7951D4E9365E CRC64;

Query Match 41.4%; Score 481.5; DB 1; Length 217;
Best Local Similarity 45.1%; Pred. No. 8.6e-34;
Matches 92; Conservative 35; Mismatches 72; Indels 5; Gaps 1.

QY 5 LGYKIKKGLVPTRLLEYLEEKYEHLERDEG----DKRRKKFELGEPNLPYTI 59
Db 4 LGYNNIRRLTAPRIIRLILEYTNNGDYERKRYNMGDAPDYSQWLNKFKLGIDFNPILPYLI 63
QY 60 DGDVLTGSMATIRIADAKHNNLGSPEKRAEISMLGCAVLDIRGVSRIVASXDFETLK 119
Db 64 DGTNRKLDGSMALIRIARIKNNLGCYTEETETIRMDLLENQYMDIRNQLMCTCSPPDEQK 123
QY 120 VDFSLKLPKMKFEDRLSHKTYLNGDHVTHPEMLYDALDVLVLYMDPCIDAFPKLYSF 179
Db 124 AEFLEGIPDKMKKLSQFGLKLPWFAGNKLTVDYDALVDLQYRMLLEPCLEAFPLKDF 183
QY 180 KKRIRAIPOIDKYLKSSKIYAMPL 203
Db 184 ISREGELEKISSYMKSRFLPKPL 207

RESULT 15
GTMU_CRILLO
ID ID GTMU_CRILLO STANDARD; PRI: 217 AA.
AC Q00285;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase Y1 (EC 2.5.1.18) (Chain 3) (GST class-mu).
OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OC NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91031445; PubMed-2226437;
RA de Saint Vincent B.R., Hyrien O., Debatisse M., Buttin G.;
RT "Complication of mu class glutathione S-transferase genes and an
  RT adenylate deaminase gene in coformycin-resistant Chinese hamster
  RT fibroblasts."
RL Eur. J. Biochem. 193:19-24(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
  OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- CATALYTIC ACTIVITY: Rx + glutathione - Hx + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.

```


Tue May 20 13:19:27 2003

us-10-081-408-5.rspt

Page 1

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 10:12:24 ; Search time 15.0413 Seconds
(Without alignments)
2986.328 Million cell updates/sec

Title: US-10-081-408-5
Perfect score: 1164
Sequence: 1 MSPILGYMKIGLVPTRL.....IAMPFGQWATFCGDPHPK 218

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1034	88.8	218	5	094745
2	726	62.4	218	5	025595
3	707	60.7	218	5	09X19
4	520.5	44.7	218	6	09N0V4
5	514.5	44.2	219	13	090MM9
6	508.5	43.7	218	11	08R516
7	508.5	43.7	223	5	097117
8	485.5	41.7	218	6	09TSM5
9	483	41.5	221	6	09BEB0
10	482.5	41.5	218	6	09TSM4
11	480.5	41.3	218	11	091Y83
12	472.5	40.6	218	11	09MU21
13	468.5	40.2	219	5	027653
14	467.5	40.2	219	5	016058
15	464.5	39.9	218	11	09DD25
16	464.5	39.9	218	11	09DCE8

17	454.5	39.0	225	4	096HA3	096ha3 homo sapien
18	448.5	38.5	225	6	09BEA9	09bea9 macaca fusc
19	444.5	38.2	195	4	005465	005465 homo sapien
20	444.5	38.2	220	4	08WME1	08wme1 homo sapien
21	442.5	38.0	232	11	09D5J8	09d5j8 mus musculu
22	440.5	37.8	225	11	09Z1B2	09z1b2 rattus norv
23	415.5	35.7	188	6	09MZB4	09mzb4 capra hircu
24	385.5	33.1	219	5	09U582	09u582 psoroptes o
25	352	30.2	181	4	08TC98	08tc98 homo sapien
26	330.5	28.4	220	5	08T7E4	08t7e4 boophilus m
27	290.5	25.0	125	6	029583	029583 sus scrofa
28	264.5	22.7	208	13	09DDU5	09ddu5 brachydanto
29	262.5	22.6	208	5	002636	002636 brugia mala
30	254.5	21.9	208	5	027711	027711 onchocerca
31	249	21.4	210	4	000460	000460 homo sapien
32	249	21.4	210	4	015680	015680 homo sapien
33	243	20.9	180	5	09NHB2	09nhb2 mytilus edu
34	242	20.8	209	6	09TY8	09ty8 capra hircu
35	222.5	19.1	216	13	09W647	09w647 oncorhynch
36	217.5	18.7	82	4	09UE37	09ue37 homo sapien
37	202.5	17.4	221	11	09DCU1	09dcu1 mus musculu
38	200.5	17.2	226	11	09JTX3	09jtx3 rattus norv
39	197.5	16.9	222	6	09N206	09n206 ovis aries
40	197	16.9	210	5	09N4X8	09n4x8 caenorhabd
41	196.5	16.9	198	5	08T1J9	08t1j9 dictyostell
42	196.5	16.9	221	11	P70686	P70686 mesocricetu
43	192.5	16.5	208	5	P91505	P91505 caenorhabd
44	191.5	16.5	223	6	029057	029057 sus scrofa
45	190.5	16.4	222	6	09X350	09x350 ovis aries

ALIGNMENTS

RESULT 1
ID 094745 PRELIMINARY; PRT; 218 AA.
AC 094745;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 26kd glutathione S-transferase.
OS Schistosoma mekongi.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeiida;
OC Schistosomatoidae; Schistosomatidae; Schistosoma.
OX NCBI_Taxid=38744;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=THAILAND;
RX MEDLINE=98222395; PubMed=9561610;
RA Grams S.V., Grams R., Korge G., Vinyant V., Upatham S.;
RT "Cloning and sequencing of the 26 kda glutathione S-transferase gene
of Schistosoma mekongi";
RL Southeast Asian J. Trop. Med. Public Health 28:570-574(1997).
DR EMBL: Y07663; CA68944.1; -.
DR HSSP: P08515; IGTA.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02796; GST_N; 1.
KW Transferase.
SQ
SEQUENCE 218 AA; 25486 MW; D3080620B19DE23A CRC64;
Query Match 88.8%; Score 1034; DB 5; Length 218;
Best Local Similarity 89.0%; Pred. No. 1.1e-81;
Matches 194; Conservative 11; Mismatches 13; Indels 0; Gaps 0;
OY 1 MSPILGYMKIGLVPTRLLEYLEEKHEHLEDECDKRNKKFELGFPNLPYID 60
DB 1 MAPILGYMKIGLVPTRLLEYLEEKHEHLEDECDKRNKKFELGFPNLPYID 60
OY 61 GDVLTOSMAITRYADKHNMLGSPKERAETISMLEGAVLDIRGVSRIVASKEFETLVK 120
|||||

Db	61	GDVLTGSMATIRYIAADKANHLGGSGKEBAEITPMLGEAVSPDIRSGVARIAYNMDFEILKV	120
Qy	121	DLSLEKLEMLKMFEDRLISHTKTYLNGDHVTHDPDLVYALDLYVLYMDPMCDADPAKLYSEK	160
Db	121	DLFKLLEMLKMFEDRLCHTKTYLNGDKVTHDPDLVYALDLYVLYMDKCKDLAPFLKYCFK	160
Qy	181	KRIEATFOIDKTYLKLSSKTYIAMPLOGNATATGGGSDHPK	218
Db	181	KRIENIPOINEYLKSSKTIEMPLDGMATATGGGSDHPK	218

RESULT 2

ID	Q25595	PRELIMINARY;	PRT;	218	AA.
AC	Q25595				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Putative glutathione transferase.				
OS	Clonorchis sinensis.				
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;				
OC	Opisthorchiida; Opisthorchiata; Opisthorchiidae; Opisthorchiidae;				
CC	Clonorchis.				
OX	NCBI_TaxID=79923;				

RP SEQUENCE FROM N.A.
RA Hong S.-J., Lee D.-H.;
RT "Cloning and over expression of 26 kDa glutathione S-transferase from
R1 *Clonorchis sinensis*.";
RL Thesis (1996), Parastology, Chung-Ang University College of Medicine,
RL Seoul.
RN 121.
RP SEQUENCE FROM N.A.
RA Hong S.-J.;
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: L47992; AAB46369.3; -
DR HSSP: P31670; 1FHE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF000043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
KW Transfeerase.
SQ 218 AA; 25038 MW; 7CBI7C7B837ADB7C CRC64;

Query Match	726	Score	726	DB	5	Length	218
Best Local Similarity	59.28	Pred. No.	4	8e-55			
Matches	129	Conservative	38	Mismatches	51	Indels	0
						Gaps	0

[illegible]

RESULTS

ID	09XYL9	PRELIMINARY;	PRT;	218 AA.
AC	09XYL9;			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DR	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DI	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE	Glutathione S-transferase.			

GN GST-1.
OS Fasciola gigantica.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomidae; Echinostomata; Fasciolidae; Fasciola
OX NCBI_TaxID=46835;
GN

RC STRAIN-THAILAND;

RA Grams S.V., Grams R., Sobhon P., Vinyant V.,
RT Molecular cloning of expressed antigens from Fasciola E.S.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBD
DR EMBL; AF12567; MAD3997.1; -
DR HSSP; P31670; 1.FHE
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PR00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
KW Transferrase.
SQ SEQUENCE 218 AA; 25337 MW; E9423D75C3F22EAF CRC64;

Query Match	60.7%	Score 707	DB 5:	length 218:
Best Local Similarity	58.7%	Pred. No.	2,1e-53:	
Matches 125;	Conservative 36;	Mismatches 52;	Indels 0;	Gaps 0

Qy	5	LGYYKINGIAGVPTLLLEYLEEKEEHEHLYEDDEDKRRNKKFELGTFPLPYIIDDVK	64
Dy	5	LGYYKINGIQPVRLLEYLEEKEEHEHLYGDRDEKRWLSDKFNMGIDLPMLPYIIDDCK	64
Qy	65	LTQSMATITRYADKHNNLGGSPKRAITSLMEGAVIDIRGVSNIAASKPEPLTKYDEL	124
Dy	65	LTQSAIDIRTYADKHGMLGSTPEERARVSMTEGGAVIDIRGVAVCYNPNEEYKGYLK	124

Qy 125 KLPEMLKMFEDRLSHKTYLNGHVTHTDFMLYALDVLVYLMDFPGLDAFPLXVSFFKRRIE 184
::: ::::
Db 125 ELPLTKMSDFLGRRQLYTSSSVSHDFMVYEALDCIRYLAPGLNDFPKLKFKFSRIE 184

```
QY 185 AIPQIDKYLSSKXIAMPLOGWQATFGGSDHP 217
      ::: ||| | |::|
Db 185 DLPKAYMESEKFIKWLPLNSWTASFGGDAP 217
```

RESULT 4

ID	ORGNOVA;	PRELIMINARY;	PRT;	218 AA.
AC	09NOVA;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)			
DE	Class mu glutathione S-transferase.			
GN	GSTM..			
OS	Bos taurus (Bovine).			
OC	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LENS.			
RA	Jimenez-sensio J.V., Garland D.;			
RT	"A lens glutathione S-transferase, class mu, with thiol-specific			
RT	antioxidant activity."			
RL	Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AF049588; AAF64308.1; -			
DR	HSSP: P03488; 1GTV			
DR	IncePro: IPR004046; GST_Cterm.			
DR	IncePro: IPR003081; GST_mu.			
DR	IncePro: IPR004045; GST_Nterm.			
DR	Pfam: PF00043; GST_C. 1.			
DR	Pfam: PF02798; GST_N. 1.			
DR	PRINTS: PRO1267; GSTNSTRASAM.			
QW	TRANSFERASE.			
SEQUENCE	218 AA; 25635 MW; 3D02EA0F43C07B0A CRC64;			

Query Match	44.7%	Score 520.5	DB 6	Length 218
Best Local Similarity	47.6%	Pred. 2.9e-37		
Matches 99	Conservative 36	Mismatches 66	Indels 5	Gaps 1

QY	1	MSPILGIWKIKGLVOPTRLLLEYEKEYEENLYERDEG----	DKVRNKKFELGLEPPPL	55
Db	1	MPMILIGWDINDIGLHAIRLLLETTDTNNEERYSGVADPDYDRSOMLNKFKLGLDFFPML		60
QY	56	PYYIDGAVKLRQSMARIYRIADKNHMLGSPKPERAEISLVEGAVLDIRGVSRIAYSKPF		115
Db	61	PYLIDGTAKTLQSNALIKYRIARKNNHLCGETEEMIRVLDLENQVADRLAMARICSPPF		120
QY	116	ETLAVDELSTKLPKMLKMFEDRLSHKTYTNGDHVTHPDEMVLDAALDVLVYMDPNCIDAPFK		175
Db	121	EKLKRGFLKEIRPEKIKLTSEFLGKRPMPWAGDKLTGYVDFLVYDVLDMHRIFFEPCIDAPFN		180
QY	176	LVSEFKKRIEATPOIDKYLKSSKYIAMP		203
Db	181	LKDFISREGKLKISATIMSSKRFLEGPL		208

RESULT 5			
090WM9			
ID	Q90WM9	PRELIMINARY;	PRT; 219 AA.
AC	Q90WM9;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Glutathione S-transferase (EC 2.5.1.18).		
GN	GSTM1.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8335;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	De Luca A., Favaloro B., Sacchetta P., Angelucci S., Di Ilio C.;		
RT	"Molecular cloning, expression and site-directed mutagenesis of a		
RT	liver Mu-class glutathione S-transferase from Xenopus laevis."		
RL	Submitted (OC7-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ146998; CAD01094.1; -.		
DR	InterPro: IPR004045; GST_Cterm.		
DR	InterPro: IPR004045; GST_Nterm.		
DR	Pfam: PF00043; GST_C; 1.		
DR	Pfam: PF02798; GST_N; 1.		
KW	Transferase.		
FT	CHAIN 2 219		XIGSTM1.
SQ	SEQUENCE 219 AA: 25359 MW: 557FFBAEDDA82DFD CRC64;		

[illegible]

RESULT
Q8R5I6

ID	QBR516	PRELIMINARY:	PRT:	218 AA.
AC	QBR516:			
DT	01-JUN-2002 (TReMBLrel. 21, Created)			
DT	01-JUN-2002 (TReMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)			
DE	Glutathione transferase GSTM7-7 (EC 2.5.1.18).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId:10090:			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-A/J:			
RA	Guo J., Zimniak L., Zimniak P., Orchard J.L., Singh S.V.;			
RT	"Cloning and expression of a novel mu class murine glutathione			
RT	transferase isoenzyme.";			
RL	Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AF464943; AAL76248.1; .			
KW	transferase			
SO	SEQUENCE	218 AA;	25519 MW;	81F15DBB46118102 CRC64;

Query Match	43.7%	Score 508.5	DB 11	Length 218
Best Local Similarity	44.9%	Pred. No. 3.2e-36		
Matches	97	Conservative	40	Mismatches 74, Indels 5, Gaps 1.

QY	1	MSPILGIYKIKGLVOPTRILLEYLEEKYBEHLERDEG----	DKRNNKFEELGLEFPNL	55
Db	1	MPMTGLGYDIRDLRAHAIIRLLLEYTGSSYEKKRYTGMADADYDRSOWLSEKFLGIDFPNL		60
QY	56	PVIIDGDVKLPDSMAIIRITADKHNHLSGSPKERAELISMLGECVAVIDIRGVSRIASKDF		115
Db	61	PLIDGSKRIITDSNAIIRKRIARKNHLGCGTEBEKIRVDLLENQANDVSNQGLRVCSPDF		120
QY	116	ETLKWDFLSKLPBEMLMKMEEDRLSHKTYTLNGDHYTHDEMFLYALDVLVLYMDPCLDAFPK		175
Db	121	EKLKLEYLEQLTGMYKVLSEFQSGORTWFEAGEITFEVDFLAYILDLHLFEPTCDAPFN		180
QY	176	LVSFKRIEALPOIDKYLLSSKYIAMPLGQWATGC	211	
Db	181	LKDFARFEVLRKISAIYMKTSRFLRPPLTKYATWG	216	

ID	097117	PRELIMINARY:	PRT:	223 AA.
AC	097117;			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)			
DE	Glutathione S-transferase.			
OS	Boophilus microplus (Cattle tick).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;			
OC	Parasitiformes; Ixodida; Ixodidae; Boophilus.			
OX	NCBI_TaxID=6941;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99381232; PubMed=10451925;			
RA	He H., Chen A.C., Davey R.B., Ivie G.W., George J.E.;			
RT	"Characterization and molecular cloning of a glutathione S-transferase			
RT	gene from the tick, Boophilus microplus (Acari: Ixodidae).";			
RL	Insect Biochem. Mol. Biol. 29:737-743(1999).			
DR	EMBL; AF077609; RAD15991.1; -.			
DR	HSSP; P20136; IGSU.			
DR	InterPro: IPR004046; GST_Cterm.			
DR	InterPro: IPR003081; GST_mu.			
DR	InterPro: IPR004045; GST_Nterm.			
DR	Pfam: PF00043; GST_C. 1.			
DR	Pfam: PF02798; GST_N. 1.			
DR	PRINTS: PR01267; GSTRNSFRASEM.			
KW	Transferase.			
SEQ	SEQUENCE 223 AA; 25604 MW; F276ACE2D8259174 CRC64;			

Query Match

43.78; Score 508.5; DB 5; Length 223;

Best Local Similarity 47.2%; Pred. No. 3.3e-36;
Matches 102; Conservative 31; Mismatches 78; Indels 5; Gaps 1;

```

OY 1 MSPITGKIKGLVPTLLLEYLEEKEEHLERD-----EGDKRNKKELEFPNL 55
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MAPVIGYMDIRGLAQPIRLILAHVADKRYTCGPPDPDRSSMLNEKTLGLEFPNL 60
OY 56 PYIDGDKVLQSMALIRYADKHNMLGSGPKERAISMELGAVLDIRYGSRIAYSKDF 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 PYIDGDKVLQSMALIRYADKHNMLGSGPKERAISMELGAVLDIRYGSRIAYSKDF 120
OY 116 ETLKVDPLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFPK 175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 ETLKVDPLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFPK 180
OY 176 LVSEFKRIEALPQIDKYLKSSKRYTAMPLOGMOATFG 211
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 LKAFYDRIEALPHVAALYKSKCIMGPLNGMVASFG 216

```

RESULT 8

```

O9TSM5 PRELIMINARY; PRT; 218 AA.
AC O9TSM5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
GN Glutathione S-transferase mu-class subunit M1 (EC 2.5.1.18).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=LIVER;
RX MEDLINE=20330602; PubMed=10869451;
RA Wang C., Bannister T.K., Guo Y., Kelly E.J., Eaton D.L.;
RT "Mu-class GSTs are responsible for aflatoxin B(1)-8, 9-epoxide-
RT conjugating activity in the nonhuman primate macaca fascicularis
RT liver.";
RL Toxicol. Sci. 56:26-36(2000).
DR EMBL; AF200709; AAF08539.1; -.
DR HSSP; P09488; 1GTU.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR003081; GST_mu.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR PRINTS; PRO1267; GSTRNSFRASEM.
KW Transferase.
SQ SEQUENCE 218 AA; 25577 MW; EE5BAD80F60C95EB CRC64;

```

Query Match 41.7%; Score 485.5; DB 6; Length 218;
Best Local Similarity 43.5%; Pred. No. 3.1e-34;
Matches 94; Conservative 37; Mismatches 80; Indels 5; Gaps 1;

```

OY 1 MSPITGKIKGLVPTLLLEYLEEKEEHLERD-----DKWRNKKELGLEFPNL 55
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MPMTLGWINDGLAHAILLEYTDSYEKKYTWGDAPDYDRSQMLNEKFXLGLDFPNL 60
OY 56 PYIDGDKVLQSMALIRYADKHNMLGSGPKERAISMELGAVLDIRYGSRIAYSKDF 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 PYIDGDKVLQSMALIRYADKHNMLGSGPKERAISMELGAVLDIRYGSRIAYSKDF 120
OY 116 ETLKVDPLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFPK 175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 ETLKVDPLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFPK 180
OY 176 LVSEFKRIEALPQIDKYLKSSKRYTAMPLOGMOATFG 211
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 LKDFISHEGLEKISAYMKSSRFPLPKPLYTRVAVWG 216

```

RESULT 9

```

O9EBE0 PRELIMINARY; PRT; 221 AA.
AC O9EBE0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
GN Glutathione transferase M2.
OS Macaca fascicularis (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9542;
RN [1]
RP SEQUENCE FROM N.A.
RA Beuckmann C.T., Fujimori K., Urade Y.;
RT "Macaca fascicularis glutathione transferase M2.";
RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB025799; BAB40442.1; -.
DR HSSP; P28161; 1HNA.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR003081; GST_mu.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR PRINTS; PRO1267; GSTRNSFRASEM.
KW Transferase.
SQ SEQUENCE 221 AA; 26068 MW; 01F22BAC6DA57FE9 CRC64;

```

Query Match 41.5%; Score 483; DB 6; Length 221;
Best Local Similarity 44.7%; Pred. No. 5.3e-34;
Matches 97; Conservative 36; Mismatches 78; Indels 6; Gaps 2;

```

OY 1 MSPITGKIKGLVPTLLLEYLEEKEEHLERD-----DKWRNKKELGLEFPNL 54
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3 MPMTLGWINDGLAHAILLEYTDSYEKKYTWGDAPDYDRSQMLNEKFXLGLDFPN 62
OY 55 LPYIDGDKVLQSMALIRYADKHNMLGSGPKERAISMELGAVLDIRYGSRIAYSKDF 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 63 LPYIDGDKVLQSMALIRYADKHNMLGSGPKERAISMELGAVLDIRYGSRIAYSKDF 122
OY 115 ETLKVDPLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFP 174
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 123 ETLKVDPLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFP 182
OY 175 LVSEFKRIEALPQIDKYLKSSKRYTAMPLOGMOATFG 211
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 183 LKDFISHEGLEKISAYMKSSRFPLPKPLYTRVAVWG 219

```

RESULT 10

```

O9TSM4 PRELIMINARY; PRT; 218 AA.
AC O9TSM4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
GN Glutathione S-transferase mu-class subunit M2 (EC 2.5.1.18).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=LIVER;
RX MEDLINE=20330602; PubMed=10869451;
RA Wang C., Bannister T.K., Guo Y., Kelly E.J., Eaton D.L.;
RT "Mu-class GSTs are responsible for aflatoxin B(1)-8, 9-epoxide-
RT conjugating activity in the nonhuman primate macaca fascicularis

```

RT liver.";
 RL Toxicol. Sci. 56:26-36(2000).
 DR EMBL: AF200710; AAF08540.1; -.
 DR HSSP: P28161; 1HNA.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTNSFRASEM.
 KM Transferrase.
 SQ SEQUENCE 218 AA; 25708 MW; F2E509C3949F9051 CRC64;

Query Match 41.5%; Score 482.5; DB 6; Length 218;
 Best Local Similarity 44.4%; Pred. No. 5,7e-34;
 Matches 96; Conservative 35; Mismatches 80; Indels 5; Gaps 1;

OY 1 MSPILGYNKIGLVOPTRLLLEYLEEKYEHLERDEG-----DKRNKKFELGLEFPNL 55
 DB 1 MPMTLGYNWINGLAHSIRLLLEYGSSYEKKRTMGADPDYDRSQWLMKEKELGLDFPNL 60
 OY 56 PYTIDGVKLTQSMALIRYADKNHMLGSPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKLTQSNALIRYARKHNLGCEETEKEKIREDLLENQMDNMQALRCYDPDF 120
 OY 116 ETLKVDLSKLPKMLKMFEDRLSHKTYLNGDVTHPDFMLYDALDVLYMDPMCLDAFPK 175
 DB 121 EKLKPEVELEGPEMLKLYSQGLKOPWFLGDKITFVDFIAYDVLRNQVFEPSCLDAFPN 180
 OY 176 LVSEKKRIEAIPOIDKYLKSSKIYAMPLOQWQATFG 211
 DB 181 LKDFISREGLKISAYMKSSRFLPRPVFTKMAVWG 216

RESULT 11
 O91Y83 PRELIMINARY; PRT; 218 AA.

AC O91Y83;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Glutathione S-transferase subunit gyc.
 OS Cavia porcellus (Guinea pig).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HARTLEY; TISSUE=LIVER;
 RX MEDLINE=98297271; PubMed=9633615;
 RA Hiratsuka A., Ogura K., Fujioke H., Sakamoto Y., Okuda H., Wada K.,
 Tanaka T., Nishiyama T., Watabe T.;
 RT "Guinea pig liver Mu-class glutathione S-transferase M1-2 cross-reacts
 with antibodies to both rat Mu- and theta-class glutathione S-
 transferases.";
 RL Arch. Biochem. Biophys. 354:188-196(1998).
 DR EMBL: AB000488; BAB47185.1; -.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferrase.
 KW Transferrase.
 SQ SEQUENCE 218 AA; 25695 MW; D2B9E31F9FEFA1E8 CRC64;

Query Match 41.7%; Score 480.5; DB 11; Length 218;
 Best Local Similarity 44.7%; Pred. No. 8,5e-34;
 Matches 93; Conservative 34; Mismatches 76; Indels 5; Gaps 1;

OY 1 MSPILGYNKIGLVOPTRLLLEYLEEKYEHLERDEG-----DKRNKKFELGLEFPNL 55
 DB 1 MPMTLGYNWINGLAHSIRLLLEYGSSYEKKRTMGADPDYDRSQWLMKEKELGLDFPNL 60
 OY 56 PYTIDGVKLTQSMALIRYADKNHMLGSPKERAISMLEGAVLDIRYGSRIAYSKDF 115

DB 61 PYLIDGTHKLTQSNALIRYARKHNLGCEETEKEKIREDLLENQMDNMQALRCYDPDF 120
 OY 116 ETLKVDLSKLPKMLKMFEDRLSHKTYLNGDVTHPDFMLYDALDVLYMDPMCLDAFPK 175
 DB 121 EKKKPEVELEGPEMLKLYSQGLKOPWFLGDKITFVDFIAYDVLRNQVFEPSCLDAFPN 180
 OY 176 LVSEKKRIEAIPOIDKYLKSSKIYAMPLOQWQATFG 203
 DB 181 LKDFISREGLKISAYMKSSRFLPRPVFTKMAVWG 208

RESULT 12
 O9WU21 PRELIMINARY; PRT; 218 AA.

AC O9WU21;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Glutathione S-transferase Yb4 (EC 2.5.1.18).
 GN GSTYB4.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RX MEDLINE=88298790; PubMed=3403534;
 RA Lai H.C., Qian B., Grove G., Tu C.P.;
 RT "Gene expression of rat glutathione S-transferases. Evidence for gene
 conversion in the evolution of the Yb multigene family.";
 RT J. Biol. Chem. 263:11389-11395(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RA Qian B., Tu C.P.;
 RT "The Rat Glutathione S-Transferase Yb4 genomic clone sequence.";
 RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF106661; AAD22630.1; -.
 DR HSSP: P04905; 2GST.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferrase.
 KW Transferrase.
 SQ SEQUENCE 218 AA; 25643 MW; 4E9C22F9AEEAC6C4 CRC64;

Query Match 40.6%; Score 472.5; DB 11; Length 218;
 Best Local Similarity 43.3%; Pred. No. 4,2e-33;
 Matches 90; Conservative 36; Mismatches 77; Indels 5; Gaps 1;

OY 1 MSPILGYNKIGLVOPTRLLLEYLEEKYEHLERDEG-----DKRNKKFELGLEFPNL 55
 DB 1 NAMITGYNVRSGLTPRILRLLEYTDSNVEEKRYWAGDAPNFRSQWLSKFLGLGDIRNL 60
 OY 56 PYTIDGVKLTQSMALIRYADKNHMLGSPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGSHKVTQSNALIRYARKHNLGCEETEKEKIREDLLENQMDNMQALRCYDPDF 120
 OY 116 ETLKVDLSKLPKMLKMFEDRLSHKTYLNGDVTHPDFMLYDALDVLYMDPMCLDAFPK 175
 DB 121 EKQKPEVELEGPEMLKLYSQGLKOPWFLGDKITFVDFIAYDVLRNQVFEPSCLDAFPN 180
 OY 176 LVSEKKRIEAIPOIDKYLKSSKIYAMPLOQWQATFG 203
 DB 181 LKDFLAREGLKISAYMKSSSFLPRPV 208

RESULT 13
 O27653 PRELIMINARY; PRT; 219 AA.
 ID O27653;
 AC O27653;

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 0.506582 seconds

(Without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30

Sequence: 1 EALFQG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	11	AA37658	Sequence encoded b
2	30	100.0	11	AA44185	C-terminus of pol
3	30	100.0	26	AA37655	Sequence encoded b
4	30	100.0	26	AA44190	N-terminus of fusi
5	30	100.0	31	AA37652	Sequence encoded b
6	30	100.0	31	AA37654	Sequence encoded b
7	30	100.0	31	AA44186	N-terminus of fusi
8	30	100.0	31	AA44189	N-terminus of fusi
9	30	100.0	989	AB92703	Herbicide-activ
10	30	100.0	2185	AA12141	Enteroviral polype

11	30	100.0	2206	13	AA22210	True type 3 poliov
12	30	100.0	2209	3	AA22037	Sequence encoded b
13	28	93.3	168	21	AA82677	Sheep erythropoiet
14	28	93.3	2164	9	AA80131	Peptides translat
15	28	93.3	2164	9	AA81045	Sequence of the vi
16	27	90.0	24	22	AA80919	Smad interacting p
17	27	90.0	76	20	AA74124	Conopeptide prop
18	27	90.0	183	22	AB87048	Drosophila melanog
19	27	90.0	333	11	AA80571	Tumour-associated
20	27	90.0	365	21	AA58734	Human FAST-1 prote
21	27	90.0	353	23	AA79681	Human protein #1 r
22	27	90.0	533	22	AB83246	Peptide #5077 enco
23	27	90.0	631	22	AA35518	Enterococcus faeca
24	27	90.0	656	23	AB85479	Lactococcus lactis
25	26	86.7	8	22	AA81065	Preselation protea
26	26	86.7	8	22	AA87810	Human Asp2 presis
27	26	86.7	8	22	AA80689	Synthetic presist
28	26	86.7	8	22	AA80656	Recognition site f
29	26	86.7	8	22	AA80725	Human Aspartyl pro
30	26	86.7	8	22	AA80264	Human Aspartyl pro
31	26	86.7	8	22	AA83541	Human rhinovirus 3
32	26	86.7	8	22	AA83054	Amino acid sequenc
33	26	86.7	8	23	AB87861	Human Asp2 related
34	26	86.7	8	23	AA87019	Human rhinovirus 3
35	26	86.7	32	16	AA85037	Peptide r1 from th
36	26	86.7	113	21	AA81498	E.coli iss protein
37	26	86.7	113	22	AA86937	Avian E coli Isola
38	26	86.7	126	22	AA89364	Human anti-Rh(D) a
39	26	86.7	127	22	AA89367	Human anti-Rh(D) a
40	26	86.7	183	22	AB85887	Drosophila pneumoni
41	26	86.7	199	20	AA34608	Chlamydia pneumonia
42	26	86.7	201	22	AA83372	Pseudomonas derugi
43	26	86.7	202	22	AA81613	Human novel secret
44	26	86.7	344	21	AA87824	D. maxinus thrombi
45	26	86.7	353	20	AA80801	Mammalian IBP-ALS

ALIGNMENTS

RESULT 1	AA37658	standard; protein; 11 AA.
ID	AA37658	
AC	AA37658	
DT	25-SEP-1993	(first entry)
XX		
DE	Sequence encoded by recombinant poliovirus at 3' end of	
DE	exogenous sequence.	
XX		
KW	Recombinant virus; proteolytic cleavage site; vaccine;	
KW	exogenous nucleic acid; replication competent.	
XX		
OS	Synthetic.	
XX		
FT	Key	Location/Qualifiers
FT	Region	1..7
FT	Cleavage-site	/label= exogenous sequence
FT	Protein	/note= "artificial"
FT		9..11
FT		/label= Polypeptide (PO)
PN	WO9311251-A.	
XX		
XX	10-JUN-1993.	
XX		
PF	04-DEC-1992.	92WO-US10543.
XX		
XX	06-DEC-1991.	91US-0804893.
PR	18-SEP-1992.	92US-0947790.
XX		

PA (AMCY) AMERICAN CYANAMID CO.
 PA (WHED) WHITEHEAD INST BIOMEDICAL.
 XX
 XX Andino R, Feinberg M, Reilly PA, Weeks-Levy CL;
 DR MPI; 1993-197068/24.
 DR N-PSDB; AAQ43325.
 XX
 PT Recombinant viruses comprising artificial proteolytic cleavage
 PT site - useful as vaccines against bacterial, viral and fungal
 PT infections, parasitic diseases, cancer and allergies
 XX
 PS Disclosure; Fig 1A; 103pp; English.
 XX
 CC There are a number of locations within the poliovirus genome at
 CC which the exogenous nucleic acid sequence encoding the exogenous
 CC polypeptide and the nucleic acid sequences encoding the artificial
 CC proteolytic cleavage sites can be positioned to produce replication-
 CC competent recombinant polioviruses that express the encoded product.
 CC These sites within the genome of the poliovirus include a terminal
 CC end, the junction between the VP1 coding region and the 2A coding
 CC region, the junction between the 2A coding region and the 2B coding
 CC region and the junction between the 2C coding region and the 3A
 CC coding region (see AAQ43394-98). AAQ43324 and AAQ43325 show poliovirus
 CC genome modifications which permit insertion of exogenous nucleic
 CC acid sequences at an end of the polioviral genome.
 CC
 SO Sequence 11 AA;
 Query Match 100.0%; Score 30; DB 14; Length 11;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 EALFQG 6
 Db 4 EALFQG 9
 RESULT 2
 ID AAY44185
 AC AAY44185 standard; peptide: 11 AA.
 AC
 AC AAY44185;
 AC
 DT 01-FEB-2000 (first entry)
 DT
 XX
 DE C-terminus of poliovirus protein for expression of exogenous genes.
 XX
 KW Antimicrobial; cytostatic; antiallergic; replication; recombinant;
 KW poliovirus; vector; expression; exogenous; protease; cleavage site;
 KW genome; primer; PCR; amplification; vaccine; bacterium; fungus; yeast;
 KW infection; parasitic disease; cancer; allergy.
 XX
 OS Synthetic.
 OS
 PN US5965124-A.
 PN
 PD 12-OCT-1999.
 PD
 PE 31-JAN-1995; 95US-0381637.
 PE
 PR 08-DEC-1992; 92US-0986729.
 PR 06-DEC-1991; 91US-0804893.
 PR 18-SEP-1992; 92US-0947790.
 PR
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA (AMCY) AMERICAN CYANAMID CO.
 PA
 PI Feinberg M, Weeks-Levy CL, Reilly PA, Andino R;
 PI
 DR MPI; 1999-633429/54.
 DR N-PSDB; AAZ30653.
 DR
 XX

PT Replication-competent recombinant poliovirus for use as vaccine against
 PT bacterial, viral, fungal, and yeast infections, parasitic diseases,
 PT cancer and allergies -
 XX
 PS Disclosure; Fig 1A; 40pp; English.
 PS
 CC The invention relates to a replication-competent recombinant poliovirus
 CC vector which is able to express an exogenous polypeptide as a component
 CC of a recombinant polyprotein precursor. The precursor molecule also
 CC contains a protease cleavage allowing the precursor to be subsequently
 CC proteolytically processed by the polio protease 3C to release the
 CC exogenous protein. Preferably the exogenous sequence and the cleavage
 CC site sequence are inserted into the vector sequence at a location in
 CC the genome of the parent poliovirus such that they do not disrupt a
 CC polioviral sequence necessary for polioviral replication. This sequence
 CC corresponds to the artificial proteolytic site which is fused to the 3'
 CC end of the exogenous gene between the exogenous protein and the fusion
 CC polypeptide from the poliovirus. The replication-competent recombinant
 CC viruses are useful as vaccines against bacterial, viral, fungal, and
 CC yeast infections, parasitic diseases, cancer and allergies.
 CC
 SO Sequence 11 AA;
 Query Match 100.0%; Score 30; DB 20; Length 11;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 EALFQG 6
 Db 4 EALFQG 9
 RESULT 3
 ID AAR37655
 AC AAR37655 standard; Protein: 26 AA.
 AC
 AC AAR37655;
 AC
 DT 25-SEP-1993 (first entry)
 DT
 XX
 DE Sequence encoded by recombinant poliovirus pMOV 3.1 (2C/3A).
 XX
 KW Recombinant virus; proteolytic cleavage site; vaccine;
 KW exogenous nucleic acid; replication competent.
 KW
 XX
 OS Synthetic.
 OS
 FH Key
 FT Cleavage-site 5..6
 FT /label- 3Cpro
 FT Cleavage-site 21..22
 FT /label- 3Cpro
 FT
 XX
 PN W09311251-A.
 PN
 PD 10-JUN-1993.
 PD
 PE 04-DEC-1992; 92WO-US10543.
 PE
 PR 06-DEC-1991; 91US-0804893.
 PR 18-SEP-1992; 92US-0947790.
 PR
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (WHED) WHITEHEAD INST BIOMEDICAL.
 PA
 PI Andino R, Feinberg M, Reilly PA, Weeks-Levy CL;
 PI
 DR MPI; 1993-197068/24.
 DR N-PSDB; AAQ43297.
 DR
 XX
 PT Recombinant viruses comprising artificial proteolytic cleavage
 PT site - useful as vaccines against bacterial, viral and fungal
 PT infections, parasitic diseases, cancer and allergies

XX Disclosure; Fig 2B; 103pp; English.

PS There are a number of locations within the poliovirus genome at

XX CC which the exogenous nucleic acid sequence encoding the exogenous

CC polypeptide and the nucleic acid sequences encoding the artificial

CC proteolytic cleavage sites can be positioned to produce replication-

CC competent recombinant polioviruses that express the encoded product.

CC These sites within the genome of the poliovirus include a terminal

CC end, the junction between the VP1 coding region and the 2A coding

CC region, the junction between the 2A coding region and the 2B coding

CC region and the junction between the 2C coding region and the 3A

CC coding region (see AAQ43294-98).

XX SQ Sequence 26 AA;

Query Match 100.0%; Score 30; DB 14; Length 26;

Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFQG 6

DB 1 EALFQG 6

RESULT 4

AA44190

ID AA44190 standard; peptide; 26 AA.

XX AC AA44190;

XX DT 01-FEB-2000 (first entry)

XX DE N-terminus of fusion protein from modified poliovirus vector pMOV3.1.

XX KM Antimicrobial; cytostatic; antiallergic; replication; recombinant;

XX KM poliovirus; vector; expression; exogenous; protease; cleavage site;

XX KM genome; primer; PCR; amplification; vaccine; bacterium; fungus; yeast;

XX KM infection; parasitic disease; cancer; allergy.

XX OS Poliovirus.

XX OS Synthetic.

XX PN US5965124-A.

XX PD 12-OCT-1999.

XX PF 31-JAN-1995; 95US-0381637.

XX PR 08-DEC-1992; 92US-0986729.

XX PR 06-DEC-1991; 91US-0804893.

XX PR 18-SEP-1992; 92US-0947790.

XX PA (MHED) WHITEHEAD INST BIOMEDICAL RES.

XX PA (AMCY) AMERICAN CYANAMID CO.

XX PI Feinberg M, Weeks-Levy CL, Reilly PA, Andino R;

XX DR WPI: 1999-632429/54.

XX DR N-PSDB: AA230658.

XX PT Replication-competent recombinant poliovirus for use as vaccine against

XX PT bacterial, viral, fungal, and yeast infections, parasitic diseases,

XX PT cancer and allergies.

XX PS Example 2; Fig 2C; 40pp; English.

XX CC The invention relates to a replication-competent recombinant poliovirus

XX CC vector which is able to express an exogenous polypeptide as a component

XX CC of a recombinant polypeptide precursor. The precursor molecule also

XX CC contains a protease cleavage allowing the precursor to be subsequently

XX CC proteolytically processed by the polio protease 3C to release the

XX CC exogenous protein. Preferably the exogenous sequence and the cleavage

CC site sequence are inserted into the vector sequence at a location in

CC the genome of the parent poliovirus such that they do not disrupt a

CC polioviral sequence necessary for polioviral replication. This sequence

CC corresponds to the N-terminal region of the fusion protein encoded by

CC the multicloning site of the poliovirus vector pMOV3.1. The sequence

CC spans the unique initiation site, the multicloning site, the protease C

CC cleavage site and part of the N-terminus of the fusion polypeptide.

CC The replication-competent recombinant viruses are useful as vaccines

CC against bacterial, viral, fungal, and yeast infections, parasitic

CC diseases, cancer and allergies.

XX SQ Sequence 26 AA;

Query Match 100.0%; Score 30; DB 20; Length 26;

Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFQG 6

DB 1 EALFQG 6

RESULT 5

AAR37652

ID AAR37652 standard; Protein; 31 AA.

XX AC AAR37652;

XX DT 25-SEP-1993 (first entry)

XX DE Sequence encoded by recombinant poliovirus pMOV 1.3 (NH2-terminal).

XX KM Recombinant virus; proteolytic cleavage site; vaccine;

XX KM exogenous nucleic acid; replication competent.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Region 16..21

FT Cleavage-site /label= poly-Gly

FT 26..27

XX FT /label= 3c pro

XX PN WO9311251-A.

XX PD 10-JUN-1993.

XX PF 04-DEC-1992; 92WO-US10543.

XX PR 06-DEC-1991; 91US-0804893.

XX PR 18-SEP-1992; 92US-0947790.

XX PA (AMCY) AMERICAN CYANAMID CO.

XX PA (MHED) WHITEHEAD INST BIOMEDICAL.

XX PI Andino R, Feinberg M, Reilly PA, Weeks-Levy CL;

XX DR WPI: 1993-197068/24.

XX DR N-PSDB: AAQ43294.

XX PT Recombinant viruses comprising artificial proteolytic cleavage

XX PT site - useful as vaccines against bacterial, viral and fungal

XX PT infections, parasitic diseases, cancer and allergies

XX PS Disclosure; Fig 2B; 103pp; English.

XX CC There are a number of locations within the poliovirus genome at

XX CC which the exogenous nucleic acid sequence encoding the exogenous

XX CC polypeptide and the nucleic acid sequences encoding the artificial

XX CC proteolytic cleavage sites can be positioned to produce replication-

XX CC competent recombinant polioviruses that express the encoded product.

XX CC These sites within the genome of the poliovirus include a terminal

XX CC end, the junction between the VP1 coding region and the 2A coding

CC region, the junction between the 2A coding region and the 2B coding
 CC region and the junction between the 2C coding region and the 3A
 CC coding region (see AAQ43294-98).

SQ Sequence 31 AA;

Query Match 100.0%; Score 30; DB 14; Length 31;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 |||||
 Db 22 EALFOG 27

RESULT 6
 AAR37654
 ID AAR37654 standard; Protein: 31 AA.

XX AAR37654;

XX 25-SEP-1993 (first entry)

XX Sequence encoded by recombinant poliovirus pmov 2.5 (2A/2B).

XX Recombinant virus: proteolytic cleavage site; vaccine;

XX exogenous nucleic acid; replication competent.

XX Synthetic.

XX Key Location/Qualifiers

FT Cleavage-site 5..6 /label- 3Cpro

FT Cleavage-site 26..27 /label- 3Cpro

XX MO9311251-A.

XX 10-JUN-1993.

XX 04-DEC-1992; 92MO-US10543.

XX 06-DEC-1991; 91US-0804893.

XX 18-SEP-1992; 92US-0947790.

XX (AMCY) AMERICAN CYANAMID CO.

XX (WHED) WHITEHEAD INST BIOMEDICAL.

XX Andino R, Feinberg M, Reilly PA, Weeks-Levy CL;

XX WPI: 1993-197068/24.

XX N-PSDB: AAQ43296.

XX Recombinant viruses comprising artificial proteolytic cleavage
 PT site - useful as vaccines against bacterial, viral and fungal
 PT infections, parasitic diseases, cancer and allergies

XX Disclosure: Fig 2B; 103pp; English.

XX There are a number of locations within the poliovirus genome at
 CC which the exogenous nucleic acid sequence encoding the exogenous
 CC polypeptide and the nucleic acid sequences encoding the artificial
 CC proteolytic cleavage sites can be positioned to produce replication-
 CC competent recombinant polioviruses that express the encoded product.
 CC These sites within the genome of the poliovirus include a terminal
 CC end, the junction between the VP1 coding region and the 2A coding
 CC region, the junction between the 2A coding region and the 2B coding
 CC coding region (see AAQ43294-98).

SQ Sequence 31 AA;

Query Match 100.0%; Score 30; DB 14; Length 31;

Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 |||||
 Db 22 EALFOG 27

RESULT 7
 AA44186
 ID AA44186 standard; peptide: 31 AA.

XX AA44186;

XX 01-FEB-2000 (first entry)

XX N-terminus of fusion protein from modified poliovirus vector pmov1.3.

XX Antimicrobial; cytostatic; antiallergic; replication; recombinant;

XX poliovirus; vector; expression; exogenous; protease; cleavage site;
 KW genome; primer; PCR; amplification; vaccine; bacterium; fungus; yeast;
 KW infection; parasitic disease; cancer; allergy.

XX Poliovirus.

XX Synthetic.

XX US5965124-A.

XX 12-OCT-1999.

XX 31-JAN-1995; 95US-0381637.

XX 08-DEC-1992; 92US-0986729.

XX 06-DEC-1991; 91US-0804893.

XX 18-SEP-1992; 92US-0947790.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX (AMCY) AMERICAN CYANAMID CO.

XX Feinberg M, Weeks-Levy CL, Reilly PA, Andino R;

XX WPI: 1999-632429/54.

XX N-PSDB: AA230634.

XX Replication-competent recombinant poliovirus for use as vaccine against
 PT bacterial, viral, fungal, and yeast infections, parasitic diseases,
 PT cancer and allergies -

XX Example 2; Fig 2B; 40pp; English.

XX The invention relates to a replication-competent recombinant poliovirus
 CC vector which is able to express an exogenous polypeptide as a component
 CC of a recombinant polypeptide precursor. The precursor molecule also
 CC contains a protease cleavage allowing the polio protease 3C to release the
 CC proteolytically processed by the polio protease 3C to release the
 CC exogenous protein. Preferably the exogenous sequence and the cleavage
 CC site sequence are inserted into the vector sequence at a location in
 CC the genome of the parent poliovirus such that they do not disrupt a
 CC polioviral sequence necessary for polioviral replication. This sequence
 CC corresponds to the N-terminal region of the fusion protein encoded by
 CC the multicloning site of the poliovirus vector pmov1.3. The sequence C
 CC spans the unique initiation site, the multicloning site, the protease C
 CC cleavage site and part of the N-terminus of the fusion polypeptide.
 CC The replication-competent recombinant viruses are useful as vaccines
 CC against bacterial, viral, fungal, and yeast infections, parasitic
 CC diseases, cancer and allergies.

SQ Sequence 31 AA;

Query Match 100.0%; Score 30; DB 20; Length 31;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
 Db 22 EALFOG 27

RESULT 8

AAV44189
 ID AAV44189 standard; peptide: 31 AA.

AC AAV44189;

DT 01-FEB-2000 (first entry)

DE N-terminus of fusion protein from modified poliovirus vector pMOV2.5.
 KM Antimicrobial; cytostatic; antiallergic; replication; recombinant;
 KM poliovirus; vector; expression; exogenous; protease; cleavage site;
 KM genome; primer; PCR; amplification; vaccine; bacterium; fungus; yeast;
 KM infection; parasitic disease; cancer; allergy.

OS Poliovirus.
 OS Synthetic.

PN USS965124-A.

PD 12-OCT-1999.

PF 31-JAN-1995; 9505-0381637.

PR 08-DEC-1992; 9205-0986729.

PR 06-DEC-1991; 9105-0804993.

PR 18-SEP-1992; 9205-0947790.

PA (WHEB) WHITEHEAD INST BIOMEDICAL RES.

PA (AMCY) AMERICAN CYANAMID CO.

PI Feinberg M, Weeks-Levy CL, Rellily PA, Andino R;

DR WPI: 1999-632429/54.

DR N-PSDB; AA220657.

XX Replication-competent recombinant poliovirus for use as vaccine against

PT bacterial, viral, fungal, and yeast infections, parasitic diseases,

PT cancer and allergies.

XX Example 2; Fig 2C; 40pp; English.

XX The invention relates to a replication-competent recombinant poliovirus
 CC vector which is able to express an exogenous polypeptide as a component
 CC of a recombinant polypeptide precursor. The precursor molecule also
 CC contains a protease cleavage allowing the precursor to be subsequently
 CC proteolytically processed by the polio protease 3C to release the
 CC exogenous protein. Preferably the exogenous sequence and the cleavage
 CC site sequence are inserted into the vector sequence at a location in
 CC the genome of the parent poliovirus such that they do not disrupt a
 CC polioviral sequence necessary for polioviral replication. This sequence
 CC corresponds to the N-terminal region of the fusion protein encoded by
 CC the multicloning site of the poliovirus vector pMOV2.5. The sequence
 CC spans the unique initiation site, the multicloning site, the protease C
 CC cleavage site and part of the N-terminus of the fusion polypeptide.
 CC The replication-competent recombinant viruses are useful as vaccines
 CC against bacterial, viral, fungal, and yeast infections, parasitic
 CC diseases, cancer and allergies.

SQ Sequence 31 AA;

Query Match 100.0%; Score 30; DB 20; Length 31;

Best Local Similarity 100.0%; Pred. No. 6.3;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
 Db 22 EALFOG 27

RESULT 9
 ABB92703
 ID ABB92703 standard; Protein: 989 AA.

AC ABB92703;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1914.

KM Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN W0200210210-A2.

PD 07-FEB-2002.

PE 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidner M;

DR WPI: 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.

PS Claim 5; SEQ ID NO 1914; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

SQ Sequence 989 AA;

Query Match 100.0%; Score 30; DB 23; Length 989;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
 Db 235 EALFOG 240

RESULT 10

AA12141
 ID AA12141 standard; Protein: 2185 AA.

AC AA12141;

DT 05-AUG-1991 (first entry)

DE Enteroviral polypeptide.

KM Enteroviruses; monoclonal antibodies; myocarditis; myositis;
 KM meningitis; encephalitis; pancreatitis; post viral fatigue.

OS Enterovirus sp.

XX

```

PN DE3939200-A.
XX
XX 29-MAY-1991.
PD
XX 27-NOV-1989; 89DE-3939200.
PF
XX 27-NOV-1989; 89DE-3939200.
PR
XX (PLAC ) MAX PLANCK GES WISSENSCH.
PA
XX Kandolf R:
PI
XX WPI; 1991-165150/23.
DR
XX N-PSDB; AA011816.
XX
XX New enteroviral polypeptide for raising group specific antibodies
PT - for detecting any type of enterovirus in blood or serum, and
PR new DNA encoding it
XX
XX Claim 1; pages 14-15; 26pp; German.
XX
XX This enteroviral polypeptide is used to raise poly- or monoclonal
CC antibodies (Abs). These are useful in assays for detecting entero-
CC virus specific antigens, as an indication of enteroviral disease.
CC All 70 serotypes of the enteroviral family can be detected.
CC Diseases such as myocarditis, myositis, meningitis, encephalitis
CC and pancreatitis can be diagnosed using the Abs.
XX
XX Sequence 2185 AA;
SO
Query Match 100.0%; Score 30; DB 12; Length 2185;
Best Local Similarity 100.0%; Pred. NO. 4.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EALFOG 6
DB 1425 EALFOG 1430
RESULT 11
AAR22210
ID AAR22210 standard; Protein; 2206 AA.
XX
XX AAR22210;
AC
XX 14-JUL-1992 (first entry)
DT
XX True type 3 poliovirus protein from LED3.
DE
XX RNA virus; error reduction.
XX
XX Poliovirus.
OS
XX MO9203538-A.
PN
XX 05-MAR-1992.
PD
XX 20-AUG-1991; 91WO-US05890.
PF
XX 20-AUG-1990; 90US-0570000.
PR 20-AUG-1990; 90US-0569916.
XX
XX (UYNV-) COLUMBIA UNIV NEW Y.
PA
XX Racanietello V, Tatem JM, Weekeslevy CL;
PI
XX WPI; 1992-096882/12.
DR
XX N-PSDB; AA022965.
XX
XX New vaccine against infectious polio-virus comprises RNA virus -
PT for producing RNA virus cDNA and viable RNA virus
PR
XX Disclosure; fig 6; 110pp; English.
PS

```

```

XX
XX The protein sequence was deduced from the cDNA sequence of P3
CC poliovirus obtd as in AA022965. The cDNA sequence is that of a
CC true RNA virus, i.e. the cDNA directs the produ. of a viable
CC RNA virus which is phenotypically similar to the source virus.
CC The full length cDNA in pLED3 was infectious. In vitro
CC transcription of pLED3 cDNA using T7 RNA polymerase produced
CC RNAs which possessed several erroneous amino acids. The RNA
CC viruses are used in vaccines against polio. The screening method
CC can be used during amplification of the source virus for vaccine
CC produ. to ensure maintenance of C at position 2493 in the viral
CC genome i.e. increasing the attenuation. The new prod. overcomes
CC the problem of errors introduced during replication of ss RNA,
CC which is much higher than for ds DNA.
XX
XX Sequence 2206 AA;
SO
Query Match 100.0%; Score 30; DB 13; Length 2206;
Best Local Similarity 100.0%; Pred. NO. 4.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EALFOG 6
DB 1449 EALFOG 1454
RESULT 12
AAP20037
ID AAP20037 standard; Protein; 2209 AA.
XX
XX AAP20037;
AC
XX 05-AUG-1992 (first entry)
DT
XX Sequence encoded by a full-length cDNA copy of the poliovirus genome
DE in plasmid pVR106.
XX
XX Poliovirus; picornavirus; vaccine; antigen; immunogen.
XX
XX Human poliovirus.
OS
XX
XX Key Location/Qualifiers
FH 1..69
FT /label= P4
FT Protein
FT 70..341
FT /label= VP2
FT Protein
FT 342..579
FT /label= VP3
FT Protein
FT 580..881
FT /label= VP1
FT Protein
FT 882..1030
FT /label= 3b
FT Protein
FT 1031..1127
FT /label= 5b
FT Protein
FT 1128..1456
FT /label= X
FT Protein
FT 1457..1543
FT /label= 1b
FT Protein
FT 1544..1565
FT /label= VPg
FT Protein
FT 1566..1748
FT /label= 2
FT Protein
FT 1749..2209
FT /label= 4(p(63))
XX
XX MO8203632-A.
PN
XX 28-OCT-1982.
PD
XX 12-NOV-1981; 81WO-3200525.
PF 12-NOV-1981; 81US-0320525.
PR 20-APR-1981; 81US-0255879.
XX

```

```

XX PA (MASI ) MASSACHUSETTS INST TECH.
XX PI
XX DR Baltimore D, Racanietello VR;
XX DR WPI; 1982-95059E/44 (95059E)..
XX DR N-PSDB; AAN20042.
XX PT Prodn. of cDNA representing viral RNA sequences - by
XX PT transcripition, insertion into vector and host cell transformation
XX PS
XX PS Example; Table 1, pages 25-31; 50pp; English.
XX CC Plasmid pVR106 was produced by combining plasmids pVR104 and pVR105.
XX CC It contains a full-length cDNA copy of the poliovirus genome. E.coli
XX CC Hst101 contg. this plasmid has been registered as ATCC 31844. The
XX CC full-length poliovirus cDNA molecule is itself infectious and can be
XX CC introduced into cells and these cultured to produce RNA virus.
XX CC Alternatively, the infectious cDNA can be treated with mutagens and
XX CC the altered material used to infect cells so that attenuated viral
XX CC RNA is prod. and this used to make vaccines. For antibody prodn.,
XX CC cDNA capable of directing antigen prodn. is selected and isolated and
XX CC incorporated into cells which are incubated to produce RNA antigen.
XX SQ
XX Sequence 2209 AA;
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 30; DB 3; Length 2209;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 EALFQG 6
XX |||||
XX Db 1452 EALFQG 1457
XX
XX RESULT 13
XX ID AAY82677 standard; Protein; 168 AA.
XX AC AAY82677;
XX DT 09-AUG-2000 (first entry)
XX DE Sheep erythropoietin protein sequence.
XX KW Erythropoietin; EPO; G-CSF; granulocyte colony stimulating factor;
XX KW wobble; codon altered gene; shuffling; modification; vaccine;
XX KW insulin; peptide hormone; growth factor; cytokine; interferon;
XX KW interleukin; leukaemia inhibitory factor; oncostatin M;
XX KW transcripition activator; expression activator; infectious organism.
XX OS
XX OS Ovis sp.
XX PN WO200018906-A2.
XX PD 06-APR-2000.
XX PF 28-SEP-1999; 99WO-US22588.
XX PR 29-SEP-1998; 98US-0102362.
XX PR 29-JAN-1999; 99US-0117729.
XX PR 05-FEB-1999; 99US-0118813.
XX PR 24-JUN-1999; 99US-0141049.
XX PA (MAXY-) MAXYGEN INC.
XX PI Patten PA, Liu L, Stemmer WPC;
XX DR WPI; 2000-303449/26.
XX PT Novel methods for recombining codon-altered libraries of nucleic acids
XX PT used to produce new proteins and new vectors with reduced rates of
XX PT reversion to wild type

```

```

XX PS Example; Fig 3; 92pp; English.
XX CC A method has been developed of making codon altered nucleic acids (Mas),
XX CC comprising providing a NA sequence (NA1) which encodes a polypeptide
XX CC (P1), providing codon altered NA sequences, each encoding P1 or a
XX CC modified form of it, and recombining the codon altered NA sequences to
XX CC produce a target codon altered NA which encodes a second protein.
XX CC The method of the invention can be used for recombining codon-altered
XX CC libraries of nucleic acids to produce new proteins, which have
XX CC improvements in a desirable characteristic. Target nucleic acids
XX CC include those coding for therapeutic proteins such as erythropoietin
XX CC (EPO), insulin, peptide hormones, growth factors, cytokines, interferons,
XX CC interleukins, leukaemia inhibitory factor, and oncostatin M, as well as
XX CC transcripition and expression activators and proteins from infectious
XX CC organisms for use as vaccines. The method can also be used to produce
XX CC attenuated viruses which have reduced rates of reversion to wild type.
XX CC The present sequence represents an EPO protein sequence, which is used
XX CC in an example from the present invention.
XX SQ
XX Sequence 168 AA;
XX
XX Query Match
XX Best Local Similarity 93.3%; Score 28; DB 21; Length 168;
XX Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 EALFQG 6
XX |||||
XX Db 72 EALFQG 77
XX
XX RESULT 14
XX ID AAP80131 standard; protein; 2164 AA.
XX AC AAP80131;
XX DT 18-OCT-1990 (first entry)
XX DE Peptides translated from Human Rhinovirus Strain 89 (HRV89) ORF.
XX KW human rhinovirus Strain 89 (HRV89); immune system stimulation; ss.
XX OS
XX OS synthetic.
XX FH Key
XX FH Location/Qualifiers
XX FT Peptide 1..69
XX FT /label=VP4
XX FT 70..336
XX FT /label=VP2
XX FT 337..574
XX FT /label=VP3
XX FT 575..872
XX FT /label=VP1
XX FT 873..1008
XX FT /label=P2-A
XX FT 1009..1103
XX FT /label=P2-B
XX FT 1104..1500
XX FT /label=P2-C
XX FT 1501..1521
XX FT /label=VPg
XX FT 1522..1704
XX FT /label=protease
XX FT 1705..3784
XX FT /label=polymerase
XX PN DE3628658-A.
XX PD 03-MAR-1988.
XX PR 23-AUG-1986; 86DE-3628658.
XX

```

```

PR 23-AUG-1986; 86DE-3628658.
XX
XX (BOEH ) BOEHRINGER INGELHEIM.
XX
XX Duchler M, Skern T, Sommergruber W, Neubauer C, Grundler P, Blaas D;
PI Kuchler E;
XX
XX WPI: 1988-064926/10.
DR N-PSDB: AAN80153.
XX
XX New DNA coding for proteins of human rhinovirus 89 -
PT and new polypeptides for stimulating immunity of blocking cell
PT receptors
XX
XX Disclosure: ; P; German.
XX
XX The viral genome is RNA with a single large open reading frame encoding
CC 10 viral peptides. These peptides can be used therapeutically to
CC stimulate the immune system in response to HRV89 and to bind to/block
CC cellular receptors.
XX
XX SQ Sequence 2164 AA;

Query Match 93.3%; Score 28; DB 9; Length 2164;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFQG 6
DB 1420 EALFQG 1425

RESULT 15
AAP81045
ID AAP81045 standard; Protein: 2164 AA.
XX
XX AAP81045;
AC
XX
XX 21-AUG-1991 (first entry)
DT
XX
DE Sequence of the viral proteins VP1-VP4, P2A-P2C, PCA-P3C encoded by
XX the genomic RNA of rhinovirus strain HRV89.
XX
XX Vaccine; diagnosis.
XX
XX Rhinovirus.
XX
XX OS
XX
XX Key Location/Qualifiers
XX FH 1.69
XX FT /label- VP4
XX FT 70..337
XX FT /label- VP2
XX FT 338..575
XX FT /label- VP3
XX FT 576..872
XX FT /label- VP1
XX FT 873..1007
XX FT /label- P2-A
XX FT 1008..1101
XX FT /label- P2-B
XX FT 1102..1421
XX FT /label- PC-2
XX FT 1422..1496
XX FT /label- P3-A
XX FT 1497..1517
XX FT /label- VPg
XX FT 1518..1678
XX FT /label- PROTEASE
XX FT 1679..2163
XX FT /label- POLYMERASE
XX
XX EP261403-A.

```

```

PD 30-MAR-1988.
XX
XX PF 20-AUG-1987; 87EP-0112104.
XX
XX PF 17-JAN-1987; 87DE-3701301.
XX
XX PR 23-AUG-1986; 86DE-3628658.
XX
XX (BOEH ) BOEHRINGER INGELHEIM.
XX
XX Duechler M, Skern T, Sommergruber W, Neubauer C, Grundler P, Blaas;
PI D, Kuchler E, Frascl L, Zorn M;
XX
XX WPI: 1988-085735/13.
DR N-PSDB: AAN81390.
XX
XX New DNA corresponding to viral RNA of rhino-virus HRV89 - useful for
PT prodn. of polypeptide(s) for stimulating immune system against HRV
PT 89.
XX
XX Clatm 8; Flg 4; 66pp; German.
XX
XX The viral proteins are used for stimulating a protective immune
CC response and for blocking cellular receptors. Abs against them
CC are useful for assay and purificn. of the corresp. antigen, and
CC can also be used for the therapeutic and diagnostic applications.
XX
XX SQ Sequence 2164 AA;

Query Match 93.3%; Score 28; DB 9; Length 2164;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFQG 6
DB 1420 EALFQG 1425

```

Search completed: May 20, 2003, 11:33:56
Job time : 1.50658 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 0.193373 Seconds
(without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30

Sequence: 1 EALFOG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCITUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/beckfilesl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	11	2	US-08-381-637-27
2	30	100.0	26	2	US-08-381-637-37
3	30	100.0	31	2	US-08-381-637-29
4	30	100.0	31	2	US-08-381-637-35
5	30	100.0	2206	1	US-07-852-260-2
6	30	100.0	2206	2	US-08-461-503-2
7	30	100.0	2206	4	US-08-465-250-2
8	27	90.0	323	6	5185254-4
9	27	90.0	365	3	US-09-113-309-2
10	27	90.0	365	4	US-09-521-109-2
11	27	90.0	365	4	US-09-562-332-2
12	26	86.7	8	4	US-09-548-372D-62
13	26	86.7	8	4	US-09-548-367D-62
14	26	86.7	32	1	US-08-190-802A-185
15	26	86.7	32	4	US-08-477-346-185
16	26	86.7	32	4	US-08-473-089-185
17	26	86.7	32	4	US-08-487-072A-185
18	26	86.7	113	3	US-09-023-221A-20
19	26	86.7	113	4	US-09-282-352A-20
20	26	86.7	126	4	US-09-240-374-152
21	26	86.7	127	4	US-09-240-374-139
22	26	86.7	353	3	US-08-986-485-6
23	26	86.7	353	3	US-08-190-802A-50
24	26	86.7	603	1	US-08-477-346-50
25	26	86.7	603	4	US-08-473-089-50
26	26	86.7	603	4	US-08-487-072A-50
27	26	86.7	1134	3	US-08-726-214-2

28	26	86.7	1568	4	US-09-181-706-2	Sequence 2, Appl1
29	26	86.7	1568	4	US-09-458-791-2	Sequence 2, Appl1
30	26	86.7	1568	4	US-09-459-066-2	Sequence 2, Appl1
31	25	83.3	19	1	US-08-383-753-63	Sequence 63, Appl1
32	25	83.3	19	2	US-08-586-772-63	Sequence 63, Appl1
33	25	83.3	19	2	US-08-959-512-63	Sequence 63, Appl1
34	25	83.3	19	4	US-09-512-983-63	Sequence 63, Appl1
35	25	83.3	21	1	US-08-828-323-21	Sequence 21, Appl1
36	25	83.3	21	4	US-08-828-323-21	Sequence 21, Appl1
37	25	83.3	178	4	US-09-315-689-5	Sequence 5, Appl1
38	25	83.3	182	4	US-09-561-500-14	Sequence 14, Appl1
39	25	83.3	182	4	US-09-561-108-14	Sequence 14, Appl1
40	25	83.3	182	4	US-09-315-689-3	Sequence 3, Appl1
41	25	83.3	182	4	US-09-561-526-14	Sequence 14, Appl1
42	25	83.3	183	4	US-09-206-059-2	Sequence 2, Appl1
43	25	83.3	184	3	US-08-741-411-1	Sequence 1, Appl1
44	25	83.3	260	3	US-09-025-059-3	Sequence 3, Appl1
45	25	83.3	260	3	US-09-070-526-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-381-637-27
Sequence 27, Application US/08381637
Patent No. 5965124
GENERAL INFORMATION:
APPLICANT: Mark Feinberg, Paul Andino, Carolyn Louise
TITLE OF INVENTION: Recombinant Vaccines and Method of
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Millita Drive
CITY: Lexington
STATE: MA
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/381,637
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/986,729
FILING DATE: 08-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH191-01AA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-381-637-27
Query Match 100.0%; Score 30; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EALFOG 6
|||||

Db 4 EALFOG 9

RESULT 2

US-08-381-637-37
Sequence 37, Application US/08381637

Patent No. 5965124

GENERAL INFORMATION:

APPLICANT: Mark Feinberg, Raul Andino, Carolyn Louise
APPLICANT: Weeks-Levy and Patricia Anne ReillyTITLE OF INVENTION: Recombinant Vaccines and Method of
NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: 2 Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381,637

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/986,729

FILING DATE: 08-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI91-01AA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-381-637-37

Query Match 100.0%; Score 30; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;OY 1 EALFOG 6
111111
Db 1 EALFOG 6

RESULT 3

US-08-381-637-29
Sequence 29, Application US/08381637

Patent No. 5965124

GENERAL INFORMATION:

APPLICANT: Mark Feinberg, Raul Andino, Carolyn Louise
APPLICANT: Weeks-Levy and Patricia Anne ReillyTITLE OF INVENTION: Recombinant Vaccines and Method of
NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: 2 Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381,637

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/986,729

FILING DATE: 08-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI91-01AA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-381-637-29

Query Match 100.0%; Score 30; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;OY 1 EALFOG 6
111111
Db 22 EALFOG 27

RESULT 4

US-08-381-637-35
Sequence 35, Application US/08381637

Patent No. 5965124

GENERAL INFORMATION:

APPLICANT: Mark Feinberg, Raul Andino, Carolyn Louise

APPLICANT: Weeks-Levy and Patricia Anne Reilly

TITLE OF INVENTION: Recombinant Vaccines and Method of

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: 2 Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381,637

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/986,729

FILING DATE: 08-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI91-01AA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-9540

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-381-637-35

Query Match 100.0%; Score 30; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 22 EALFOG 27

RESULT 5
US-07-852-260-2
Sequence 2, Application US/07852260
Patent No. 5525715
GENERAL INFORMATION:
APPLICANT: Racanelli, Vincent
APPLICANT: Tatem, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/852,260
FILING DATE: 19920619
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36607-B-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2206 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-852-260-2

Query Match 100.0%; Score 30; DB 1; Length 2206;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1449 EALFOG 1454

RESULT 6
US-08-461-503-2
Sequence 2, Application US/08461503
Patent No. 5834302
GENERAL INFORMATION:

APPLICANT: Racanelli, Vincent
APPLICANT: Tatem, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES
TITLE OF INVENTION: FROM CDNA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,503
FILING DATE: 5-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36607-D-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2206 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-461-503-2

Query Match 100.0%; Score 30; DB 2; Length 2206;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1449 EALFOG 1454

RESULT 7
US-08-465-250-2
Sequence 2, Application US/08465250
Patent No. 6136570
GENERAL INFORMATION:
APPLICANT: Racanelli, Vincent
APPLICANT: Tatem, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release 1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,250
FILING DATE: 6-JUN-1995

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 36607-E-PCT-US
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2206 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-250-2

Query Match
Best Local Similarity 100.0%; Score 30; DB 4; Length 2206;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1449 EALFOG 1454

RESULT 8
; Patent No. 5185254
; APPLICANT: LINNENBACH, ALBAN
; TITLE OF INVENTION: GENE FAMILY OF TUMOR-ASSOCIATED ANTIGENS
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION NUMBER:
; APPLICATION NUMBER: US/07/291,583
; FILING DATE: 29-DEC-1988
; SEQ ID NO:4:
; LENGTH: 323
5185254-4

Query Match
Best Local Similarity 90.0%; Score 27; DB 6; Length 323;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 233 EALFOG 238

RESULT 9
US-09-113-309-2
; Sequence 2, Application US/09113309A
; Patent No. 6110738
; GENERAL INFORMATION:
; APPLICANT: Zhou, Shihlin
; APPLICANT: Zawal, Leigh
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Human Fast-1 Gene
; FILE REFERENCE: 01107,10898
; CURRENT APPLICATION NUMBER: US/09/113,309A
; CURRENT FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-113-309-2

Query Match
Best Local Similarity 90.0%; Score 27; DB 3; Length 365;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 EALFOG 6
Db 328 DALFOG 333

RESULT 10
US-09-521-109-2
; Sequence 2, Application US/09521109
; Patent No. 6225441
; GENERAL INFORMATION:
; APPLICANT: Zhou, Shihlin
; APPLICANT: Zawal, Leigh
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Human Fast-1 Gene
; FILE REFERENCE: 01107,10898
; CURRENT APPLICATION NUMBER: US/09/521,109
; CURRENT FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 09/113,309
; PRIOR FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-521-109-2

Query Match
Best Local Similarity 90.0%; Score 27; DB 4; Length 365;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 328 DALFOG 333

RESULT 11
US-09-562-332-2
; Sequence 2, Application US/09562332
; Patent No. 6372434
; GENERAL INFORMATION:
; APPLICANT: Zhou, Shihlin
; APPLICANT: Zawal, Leigh
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Human Fast-1 Gene
; FILE REFERENCE: 01107,10898
; CURRENT APPLICATION NUMBER: US/09/562,332
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/113,309
; PRIOR FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-562-332-2

Query Match
Best Local Similarity 90.0%; Score 27; DB 4; Length 365;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 328 DALFOG 333

RESULT 12
US-09-548-372D-62
; Sequence 62, Application US/09548372D
; Patent No. 6420534
```


GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 29915/62801
; CURRENT APPLICATION NUMBER: US/09/548,372D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-548-372D-62

Query Match 86.7%; Score 26; DB 4; Length 8;
Best Local Similarity 83.3%; Pred. No. 1.9e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EALFOG 6
1
1
1
1
1
DB 2 EYLFOG 7

RESULT 13
US-09-548-367D-62
; Sequence 62, Application US/09548367D
; Patent No. 6440698
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 29915/6280H
; CURRENT APPLICATION NUMBER: US/09/548,367D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-548-367D-62

Query Match 86.7%; Score 26; DB 4; Length 8;
Best Local Similarity 83.3%; Pred. No. 1.9e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EALFOG 6
1
1
1
1
1
DB 2 EYLFOG 7

RESULT 14

US-08-190-802A-185
; Sequence 185, Application US/08190802A
; Patent No. 5519003
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,802A
; FILING DATE: 01-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Insulin-like growth factor bind.
; INDIVIDUAL ISOLATE: pro. complex-rat r1, Fig. 33
US-08-190-802A-185

Query Match 86.7%; Score 26; DB 1; Length 32;
Best Local Similarity 83.3%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EALFOG 6
1
1
1
1
1
DB 21 EGLFOG 26

RESULT 15
US-08-477-346-185
; Sequence 185, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Insulin-like growth factor bind.
INDIVIDUAL ISOLATE: pro. complex-rat II, Fig. 33
US-08-477-346-185

Query Match 86.7%; Score 26; DB 4; Length 32;
Best Local Similarity 83.3%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EALFQG 6
| | | | |
DB 21 EGLFQG 26

Search completed: May 20, 2003, 11:40:58
Job time: 1.19337 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 0.955969 Seconds

(without alignments)
622.443 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30

Sequence: 1.EALFGG 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	6	9 US-10-081-408-6	Sequence 5, Appli
2	30	100.0	998	9 US-10-081-408-20	Sequence 20, Appli
3	27	90.0	323	9 US-10-097-340-314	Sequence 314, App
4	27	90.0	533	10 US-09-864-761-48941	Sequence 48941, A
5	27	90.0	631	10 US-09-815-242-10711	Sequence 10711, A
6	27	90.0	707	9 US-09-884-696-21	Sequence 21, Appli
7	26	86.7	8	9 US-09-869-414-62	Sequence 62, Appli
8	26	86.7	8	10 US-09-794-927-62	Sequence 62, Appli
9	26	86.7	8	10 US-09-795-847-62	Sequence 62, Appli
10	26	86.7	8	10 US-09-794-743-62	Sequence 62, Appli
11	26	86.7	8	10 US-09-794-743-62	Sequence 62, Appli
12	26	86.7	8	10 US-09-794-743-62	Sequence 62, Appli
13	26	86.7	8	10 US-09-794-925-62	Sequence 62, Appli
14	26	86.7	14	9 US-10-057-789-63	Sequence 62, Appli
15	26	86.7	14	9 US-10-057-789-63	Sequence 62, Appli
16	26	86.7	14	9 US-10-212-628-63	Sequence 64, Appli
17	26	86.7	14	9 US-10-212-628-64	Sequence 64, Appli
18	26	86.7	20	9 US-10-057-789-61	Sequence 61, Appli
19	26	86.7	20	9 US-10-057-789-62	Sequence 62, Appli

20	26	86.7	20	9 US-10-212-628-61	Sequence 61, Appli
21	26	86.7	20	9 US-10-212-628-62	Sequence 62, Appli
22	26	86.7	113	10 US-09-738-559-70	Sequence 20, Appli
23	26	86.7	126	9 US-09-848-798-152	Sequence 152, App
24	26	86.7	127	9 US-09-848-798-152	Sequence 152, App
25	26	86.7	201	10 US-09-815-242-5217	Sequence 5217, Ap
26	26	86.7	202	10 US-09-764-864-1126	Sequence 1126, Ap
27	26	86.7	1024	10 US-09-864-921-97	Sequence 97, Appli
28	26	86.7	1024	10 US-09-841-739-2	Sequence 2, Appli
29	26	86.7	1204	10 US-09-841-739-5	Sequence 5, Appli
30	25	83.3	9	10 US-09-786-284-45	Sequence 21, Appli
31	25	83.3	9	10 US-09-786-284-45	Sequence 21, Appli
32	25	83.3	11	9 US-09-879-572A-23	Sequence 23, Appli
33	25	83.3	11	9 US-09-879-572A-25	Sequence 25, Appli
34	25	83.3	50	9 US-09-879-572A-32	Sequence 32, Appli
35	25	83.3	63	10 US-09-832-540A-1	Sequence 1, Appli
36	25	83.3	77	9 US-09-879-572A-30	Sequence 30, Appli
37	25	83.3	96	10 US-09-864-761-34961	Sequence 34961, A
38	25	83.3	132	9 US-09-764-891-4364	Sequence 4364, Ap
39	25	83.3	178	9 US-10-131-241-60	Sequence 60, Appli
40	25	83.3	179	9 US-10-131-241-57	Sequence 57, Appli
41	25	83.3	180	9 US-10-131-241-55	Sequence 56, Appli
42	25	83.3	181	9 US-10-131-241-55	Sequence 55, Appli
43	25	83.3	182	9 US-10-131-241-54	Sequence 54, Appli
44	25	83.3	182	10 US-09-998-831-14	Sequence 14, Appli
45	25	83.3	183	9 US-10-080-797-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-081-408-6
Sequence 6, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abrahms, n , Lars
APPLICANT: Nilsson, Joakim
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 6
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Protease cleavage site
US-10-081-408-6

Query Match 100.0%; Score 30; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 3.3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;

Db 1 EALFGG 6
OY 1 EALFGG 6
|||||
Db 1 EALFGG 6

RESULT 2
US-10-081-408-20
Sequence 20, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abrahms, n , Lars
APPLICANT: Nilsson, Joakim
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION

FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: US 60/272,247
PRIOR FILING DATE: 2001-02-28
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 20
LENGTH: 998
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant construct
US-10-081-408-20

Query Match 100.0%; Score 30; DB 9; Length 998;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EALFOG 6
Db 259 EALFOG 264

RESULT 3
US-10-097-340-314
Sequence 314, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:
APPLICANT: Manjula GANNAVAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATIS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIRY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, JR.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumel ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
Title Of Invention: Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097,340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 314
LENGTH: 323
TYPE: PRT
ORGANISM: Homo sapiens
US-10-097-340-314

Query Match 90.0%; Score 27; DB 9; Length 323;
Best Local Similarity 83.3%; Pred. No. 1,4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EALFOG 6
Db 233 EALFOG 238

RESULT 4
US-09-864-761-48941
Sequence 48941, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-06-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 48941
LENGTH: 533
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008174.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.63
OTHER INFORMATION: EST_HUMAN HIT: BF129490.1, EVALUATE 1.30e+00
OTHER INFORMATION: SWISSPROT HIT: P13666, EVALUATE 6.40e+02
US-09-864-761-48941

Query Match 90.0%; Score 27; DB 10; Length 533;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
11:1111
DB 494 EALFOG 499

RESULT 5
US-09-815-242-10711

; Sequence 10711, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trivick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10711
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10711

Query Match 90.0%; Score 27; DB 10; Length 631;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
11:1111
DB 89 EALFOG 94

RESULT 6
US-09-884-696-21

; Sequence 21, Application US/09884696
; Publication No. US20030035809A1
; GENERAL INFORMATION:
; APPLICANT: GEORGE, Lisle W
; APPLICANT: ANGELIOS, JOHN A
; APPLICANT: HESS, JOHN F
; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES
; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA
; TITLE OF INVENTION: BOVIS INJECTIONS
; FILE REFERENCE: 481.06
; CURRENT APPLICATION NUMBER: US/09/884,696
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21

; LENGTH: 707

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-884-696-21

Query Match 90.0%; Score 27; DB 9; Length 707;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
11:1111
DB 115 EALFOG 120

RESULT 7
US-09-869-414-62

; Sequence 62, Application US/09869414
; Publication No. US20030077226A1
; GENERAL INFORMATION:
; APPLICANT: Belinkowski et al.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 28341/6280X
; CURRENT APPLICATION NUMBER: US/09/869,414
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-869-414-62

Query Match 86.7%; Score 26; DB 9; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
11:1111
DB 2 EALFOG 7

RESULT 8
US-09-794-927-62

; Sequence 62, Application US/09794927
; Patent No. US20010016324A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Helinikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Van Riddling
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 28341/6280FG
; CURRENT APPLICATION NUMBER: US/09/794,927
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/416,901

```
;; PRIOR FILING DATE: 1999-10-13
;; PRIOR APPLICATION NUMBER: 60/155,493
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: 09/404,133
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: PCT/US99/20881
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: 60/101,594
;; PRIOR FILING DATE: 1998-09-24
;; NUMBER OF SEQ ID NOS: 73
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 62
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-794-927-62
```

```
Query Match      86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 EALFQG 6
        | | | | |
Db       2 EVLFGG 7
```

```
RESULT 9
US-09-795-847-62
;; Sequence 62, Application US/09795847
;; Patent No. US20010018208A1
;; GENERAL INFORMATION:
;; APPLICANT: Gurney, Mark E.
;; APPLICANT: Bienkowski, Michael J.
;; APPLICANT: Heinrichson, Robert L.
;; APPLICANT: Parodi, Luis A.
;; APPLICANT: Yan, Riqiang
;; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
;; TITLE OF INVENTION: USES
;; FILE REFERENCE: 28341/6280DE
;; CURRENT APPLICATION NUMBER: US/09/795,847
;; CURRENT FILING DATE: 2001-02-28
;; PRIOR APPLICATION NUMBER: 09/416,901
;; PRIOR FILING DATE: 1999-10-13
;; PRIOR APPLICATION NUMBER: 60/155,493
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: 09/404,133
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: PCT/US99/20881
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: 60/101,594
;; PRIOR FILING DATE: 1998-09-24
;; NUMBER OF SEQ ID NOS: 73
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 62
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-795-847-62
```

```
Query Match      86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 EALFQG 6
        | | | | |
Db       2 EVLFGG 7
```

```
RESULT 10
US-09-794-743-62
;; Sequence 62, Application US/09794743
;; Patent No. US20010021391A1
;; GENERAL INFORMATION:
;; APPLICANT: Gurney, Mark E.
;; APPLICANT: Bienkowski, Michael J.
;; APPLICANT: Heinrichson, Robert L.
;; APPLICANT: Parodi, Luis A.
;; APPLICANT: Yan, Riqiang
;; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
;; TITLE OF INVENTION: USES
;; FILE REFERENCE: 28341/6280BC
;; CURRENT APPLICATION NUMBER: US/09/794,743
;; CURRENT FILING DATE: 2001-02-27
;; PRIOR APPLICATION NUMBER: 09/416,901
;; PRIOR FILING DATE: 1999-10-13
;; PRIOR APPLICATION NUMBER: 60/155,493
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: 09/404,133
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: PCT/US99/20881
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: 60/101,594
;; PRIOR FILING DATE: 1998-09-24
;; NUMBER OF SEQ ID NOS: 73
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 62
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-794-743-62
```

```
Query Match      86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 EALFQG 6
        | | | | |
Db       2 EVLFGG 7
```

```
RESULT 11
US-09-794-748-62
;; Sequence 62, Application US/09794748
;; Patent No. US20020037315A1
;; GENERAL INFORMATION:
;; APPLICANT: Gurney, Mark E.
;; APPLICANT: Bienkowski, Michael J.
;; APPLICANT: Heinrichson, Robert L.
;; APPLICANT: Parodi, Luis A.
;; APPLICANT: Yan, Riqiang
;; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
;; TITLE OF INVENTION: USES
;; FILE REFERENCE: 28341/6280JL
;; CURRENT APPLICATION NUMBER: US/09/794,748
;; CURRENT FILING DATE: 2001-02-27
;; PRIOR APPLICATION NUMBER: 09/416,901
;; PRIOR FILING DATE: 1999-10-13
;; PRIOR APPLICATION NUMBER: 60/155,493
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: 09/404,133
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: PCT/US99/20881
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: 60/101,594
;; PRIOR FILING DATE: 1998-09-24
;; NUMBER OF SEQ ID NOS: 73
;; SOFTWARE: Patentln Ver. 2.0
```

SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-794-748-62

Query Match
Best Local Similarity 86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1 ||||
Db 2 EVLFOG 7

RESULT 12
US-09-794-925-62
Sequence 62, Application US/09794925
Patent No. US20020064819A1
GENERAL INFORMATION:
APPLICANT: Gurney, Mark E.
APPLICANT: Bienkowski, Michael J.
APPLICANT: Heinrikson, Robert L.
APPLICANT: Parodi, Luis A.
APPLICANT: Yan, Riqiang
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND USES
FILE REFERENCE: 28341/6280H1
CURRENT FILING DATE: 2001-02-27
CURRENT APPLICATION NUMBER: US/09/794,925
PRIOR FILING DATE: 1999-10-13
PRIOR APPLICATION NUMBER: 60/155,493
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 09/404,133
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: PCT/US99/20881
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 60/101,594
PRIOR FILING DATE: 1998-09-24
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-794-925-62

Query Match
Best Local Similarity 86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1 ||||
Db 2 EVLFOG 7

RESULT 13
US-09-681-442-62
Sequence 62, Application US/09681442
Patent No. US20020081634A1
GENERAL INFORMATION:
APPLICANT: Gurney, Mark E.
APPLICANT: Bienkowski, Michael J.
APPLICANT: Heinrikson, Robert L.
APPLICANT: Parodi, Luis A.
APPLICANT: Yan, Riqiang
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND USES
TITLE OF INVENTION: THEREFOR

FILE REFERENCE: 28341/6280H1
CURRENT APPLICATION NUMBER: US/09/681,442
CURRENT FILING DATE: 2001-04-05
PRIOR FILING DATE: 1999-10-13
PRIOR APPLICATION NUMBER: 60/155,493
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 09/404,133
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: PCT/US99/20881
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 60/101,594
PRIOR FILING DATE: 1998-09-24
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-681-442-62

Query Match
Best Local Similarity 86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1 ||||
Db 2 EVLFOG 7

RESULT 14
US-10-057-789-63
Sequence 63, Application US/10057789
Publication No. US20030082522A1
GENERAL INFORMATION:
APPLICANT: Paul Haynes
APPLICANT: Jing Wei
APPLICANT: John Yates
APPLICANT: Nancy Andon
TITLE OF INVENTION: ANALYSIS OF COMPLEX PROTEIN MIXTURES
FILE REFERENCE: NADIT-022A
CURRENT APPLICATION NUMBER: US/10/057,789
CURRENT FILING DATE: 2002-06-28
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/264,576
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 60/305,232
NUMBER OF SEQ ID NOS: 311
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 63
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthesized Peptide
US-10-057-789-63

Query Match
Best Local Similarity 86.7%; Score 26; DB 9; Length 14;
Best Local Similarity 83.3%; Pred. No. 8;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1 ||||
Db 7 EVLFOG 12

RESULT 15
US-10-057-789-64
Sequence 64, Application US/10057789
Publication No. US20030082522A1

GENERAL INFORMATION:
APPLICANT: Paul Haynes
APPLICANT: Jing Wei
APPLICANT: John Yates
APPLICANT: Nancy Andon
TITLE OF INVENTION: DIFFERENTIAL LABELING FOR QUANTITATIVE
FILE REFERENCE: NADII.022A
CURRENT APPLICATION NUMBER: US/10/057,789
CURRENT FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: US 60/264,576
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/305,232
PRIOR FILING DATE: 2001-07-13
NUMBER OF SEQ ID NOS: 311
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 64
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthesized Peptide
FEATURE:
NAME/KEY: VARIANT
LOCATION: 14
OTHER INFORMATION: Xaa - Ornithine
US-10-057-789-64

Query Match 86.7%; Score 26; DB 9; Length 14;
Best Local Similarity 83.3%; Pred. NO. 8;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EALFQG 6
Db 7 EVLFQG 12

Search completed: May 20, 2003, 12:02:22
Job time : 0.955969 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:09:54 ; Search time 0.231502 Seconds

(without alignments)
2491.578 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30
Sequence: 1 EALFOG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	989	2 T46183	zinc proteinase (E
2	30	100.0	1081	2 S15040	pleiotropic drug r
3	30	100.0	1374	1 GNNYEC	genome polypotein
4	30	100.0	2175	1 GNNYBE	genome polypotein
5	30	100.0	2182	1 GNNYB1	genome polypotein
6	30	100.0	2183	1 GNNYB4	genome polypotein
7	30	100.0	2185	1 GNNYB3	genome polypotein
8	30	100.0	2185	1 GNNYBT	genome polypotein
9	30	100.0	2185	1 JQ2021	genome polypotein
10	30	100.0	2185	1 GNNYSV	genome polypotein
11	30	100.0	2185	1 GNNYSV	genome polypotein
12	30	100.0	2193	2 S44194	genome polypotein
13	30	100.0	2193	2 S44194	genome polypotein
14	30	100.0	2193	2 S52919	polyprotein - echo
15	30	100.0	2194	2 S52919	polyprotein (1A, 1
16	30	100.0	2201	1 GNNYF7	genome polypotein
17	30	100.0	2201	1 GNNYF9	genome polypotein
18	30	100.0	2206	1 GNNY4P	genome polypotein
19	30	100.0	2206	1 GNNY27	genome polypotein
20	30	100.0	2206	2 S03822	genome polypotein
21	30	100.0	2207	1 GNNY1P	genome polypotein
22	30	100.0	2207	1 GNNY1P	genome polypotein
23	30	100.0	2207	1 GNNY1P	genome polypotein
24	30	100.0	2209	1 GNNY2P	genome polypotein
25	30	100.0	2209	1 GNNY3P	genome polypotein
26	30	100.0	2214	1 A48548	genome polypotein
27	28	93.3	247	2 A87511	conserved hypotet
28	28	93.3	278	2 E83152	probable short-cha
29	28	93.3	425	2 T10702	phycocene synthase

30	28	93.3	445	2 AG1556	Na ⁺ -transporting A
31	28	93.3	2164	1 GNNY89	genome polypotein
32	27	90.0	117	2 A95907	probable protein i
33	27	90.0	117	2 G95924	hypothetical prote
34	27	90.0	201	2 A82466	hypothetical prote
35	27	90.0	227	2 AB2008	hypothetical prote
36	27	90.0	256	2 T36793	probable oxidoredu
37	27	90.0	272	2 S57460	cytochrome-c oxida
38	27	90.0	323	1 A48149	carcinoma-associat
39	27	90.0	338	2 A82209	conserved hypotet
40	27	90.0	380	2 T40541	hypothetical prote
41	27	90.0	395	2 H70385	tryptophan-tRNA 11
42	27	90.0	442	2 T18917	hypothetical prote
43	27	90.0	443	2 T11804	translation initia
44	27	90.0	452	2 S22199	imidazoleglycerol-
45	27	90.0	614	2 B43935	spat protein - Bac

ALIGNMENTS

RESULT 1
T46183
zinc proteinase (EC 3.4.24.-) T8H10.70 - Arabidopsis thaliana
N:Alternate names: Insulysin homolog
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 17-Mar-2000
C:Accession: T46183
R:Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223014
A:Accession: T46183
A:Molecule type: DNA
A:Residues: 1-989 <BEN>
A:Cross-References: EMBL:AL133248; PIDN:CAB6104.1
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:introns: 48/3: 74/1; 96/3: 127/2; 152/1; 166/3; 184/1; 225/1; 262/3; 317/1; 348/1
A:Note: T8H10.70
C:Superfamily: Insulysin
C:Keywords: hydrolase; metalloproteinase; zinc
F:1/75/Binding site: zinc (His) #status predicted
F:74/Active site: Glu #status predicted

Query Match 100.0%; Score 30; DB 2; Length 989;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 235 EALFOG 240
RESULT 2
S15040
pleiotropic drug resistance control protein PR6 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G313; protein YGL016w
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 21-Jul-2000
R:Chen, W.; Balz, E.; Capleau, E.; Choder, M.; Goffeau, A.
Yeast 7, 287-299, 1991
A:Title: The DNA sequencing of the 17 kb HindIII fragment spanning the LEU1 and ATE
ntroiling pleiotropic drug resistance.
A:Reference number: S15040; MUID:91353083; PMID:1882553
A:Accession: S15040
A:Molecule type: DNA
A:Residues: 1-1081 <CHE>
A:Cross-References: GB:S57895; NID:9234319; PIDN:AAB19613.1; PID:9234320
R:Hebling, U.; Hofmann, B.; Delius, H.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64003

A.Accession: S64018
 A.Molecule type: DNA
 A.Residues: 1-1081 <HEB>
 A.Cross-references: EMBL:272538; NID:q1322476; PID:6243847; PID:q1322477; MIPS:YGL016w
 A.Experimental source: strain S288C
 C.Genetics:
 A.Gene: SGD:PD86
 A.Cross-references: SGD:S0002984; MIPS:YGL016w
 A.Map position: 7L
 C.Keywords: transmembrane protein
 F:355-381/Domain: transmembrane #status predicted <TM2>
 F:540-556/Domain: transmembrane #status predicted

Query Match 100.0%; Score 30; DB 2; Length 1081;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 631 EALFOG 636

RESULT 3

GNNYFC

genome polypotein - echovirus 11 (strain Gregory) (fragment)
 N:Contains: carboxyl end of coat protein 1D; core protein 2A; core protein 2B; core prot
 C:Species: echovirus 11
 C.Date: 31-Dec-1992 #sequence #revision 31-Dec-1992 #text_change 16-Jun-2000
 C.Accession: A36642
 R:Auvinen, P.; Hyypia, T.
 J.Gen. Virol. 71, 2133-2139, 1990
 A.Title: Echoviruses include genetically distinct serotypes.
 A.Reference number: A36642; MUID:91011360; PMID:2170575
 A.Accession: A36642
 A.Molecule type: genomic RNA
 A.Residues: 1-1374 <AU>
 A.Cross-references: GB:D10582; GB:D01068; NID:g465394; PIDN:BA01439.1; PID:g465395
 C:Superfamily: poliovirus genome polypotein
 C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase; F
 F:1-43/Product: coat protein 1D (fragment) #status predicted <VP1>
 F:44-190/Product: core protein 2A #status predicted <2AP>
 F:191-289/Product: core protein 2B #status predicted <2BP>
 F:290-618/Product: core protein 2C #status predicted <2CP>
 F:619-707/Product: core protein 3A #status predicted <P3A>
 F:708-729/Product: genome-linked protein VPg #status predicted <VPg>
 F:730-912/Product: proteinase #status predicted <P3S>
 F:913-1374/Product: RNA-directed RNA polymerase #status predicted <RPS>
 F:710/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 1374;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 614 EALFOG 619

RESULT 4

GNNYBE

genome polypotein - bovine enterovirus (strain VG-5-27)
 N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core
 RNA-directed RNA polymerase (EC 2.7.7.48)
 C:Species: bovine enterovirus
 C.Date: 30-Jun-1989 #sequence #revision 30-Jun-1989 #text_change 16-Jun-2000
 C.Accession: A29824
 R:Earle, J.A.P.; Skuce, R.A.; Fleming, C.S.; Hoey, E.M.; Martin, S.J.
 J.Gen. Virol. 69, 253-263, 1988
 A.Title: The complete nucleotide sequence of a bovine enterovirus.
 A.Reference number: A29824; MUID:88117392; PMID:2828511
 A.Accession: A29824
 A.Molecule type: genomic RNA
 A.Residues: 1-2175 <EAR>

A.Cross-references: GB:D00214; NID:g2696866; PIDN:BA024003.1
 C:Superfamily: poliovirus genome polypotein
 C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase
 F:1-69/Product: coat protein VP4 #status predicted <VP4>
 F:70-317/Product: coat protein VP2 #status predicted <VP2>
 F:318-558/Product: coat protein VP3 #status predicted <VP3>
 F:560-840/Product: coat protein VP1 #status predicted <VP1>
 F:841-990/Product: core protein P2-A #status predicted <P2A>
 F:991-1089/Product: core protein P2-B #status predicted <P2B>
 F:1090-1419/Product: core protein P2-C #status predicted <P2C>
 F:1420-1508/Product: core protein P3-A #status predicted <P3A>
 F:1509-1531/Product: genome-linked protein VPg #status predicted <VPg>
 F:1532-1714/Product: proteinase #status predicted <P3S>
 F:1715-2175/Product: RNA-directed RNA polymerase #status predicted <RPS>
 F:1511/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2175;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1415 EALFOG 1420

RESULT 5

GNNYB1

genome polypotein - coxsackievirus B1
 N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
 polymerase (EC 2.7.7.48)
 C:Species: coxsackievirus B1
 C.Date: 30-Sep-1988 #sequence #revision 30-Sep-1988 #text_change 16-Jul-1999
 C.Accession: A26353
 R:Iizuka, N.; Kuge, S.; Nomoto, A.
 Virol. 156, 64-73, 1987
 A.Title: Complete nucleotide sequence of the genome of coxsackievirus B1.
 A.Reference number: A26353; MUID:87122157; PMID:3027969
 A.Accession: A26353
 A.Molecule type: genomic RNA
 A.Residues: 1-2182 <IIZ>
 A.Cross-references: GB:M6560; NID:g323417; PIDN:AA00531.1; PID:g323418
 C:Superfamily: poliovirus genome polypotein
 C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase
 F:1-69/Product: coat protein 1A #status predicted <P1A>
 F:70-332/Product: coat protein 1B #status predicted <P1B>
 F:333-570/Product: coat protein 1C #status predicted <P1C>
 F:571-848/Product: coat protein 1D #status predicted <P1D>
 F:849-998/Product: core protein 2A #status predicted <C2A>
 F:999-1097/Product: core protein 2B #status predicted <C2B>
 F:1098-1426/Product: core protein 2C #status predicted <C2C>
 F:1427-1515/Product: protein 3A #status predicted <P3A>
 F:1516-1537/Product: genome-linked protein VPg #status predicted <VPg>
 F:1538-1720/Product: proteinase #status predicted <P3S>
 F:1721-2182/Product: RNA-directed RNA polymerase #status predicted <RNS>
 F:1516/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2182;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1422 EALFOG 1427

RESULT 6

GNNYB4

genome polypotein - coxsackievirus B4
 N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
 polymerase (EC 2.7.7.48)
 C:Species: coxsackievirus B4
 C.Date: 31-Mar-1989 #sequence #revision 31-Mar-1989 #text_change 16-Jul-1999
 C.Accession: A27170; A45576; B45576

R.Jenkins, O.; Booth, J.D.; Minor, P.D.; Almond, J.W.
J. Gen. Virol. 68, 1835-1848, 1987
A:Title: The complete nucleotide sequence of coxsackievirus B4 and its comparison to oth
A:Reference number: A27170; MUID:87253111; PMID:3037008
A:Accession: A27170
A:Molecule type: genomic RNA
A:Residues: 1-2183 <JEN>
R.Ramalingh, A.; Araki, H.; Bryant, S.; Hixson, A.
Virus Res. 23, 281-292, 1992
A:Title: Identification of candidate sequences that determine virulence in Cocksackievir
A:Reference number: A45576; MUID:92327853; PMID:1320798
A:Accession: A45576
A:Molecule type: genomic RNA
A:Residues: 569-696; 'T', 698-852 <RAM>
A:Cross-references: GB:S39291; NID:9250908; PIDN:AA82445.1; PID:9250909
A:Experimental source: B4; virulent strain
A:Note: sequence inconsistent with the nucleotide translation
A:Accession: B45576
A:Molecule type: genomic RNA
A:Residues: 70-203; 'A', 205-231; 'E', 233-269; 'A', 271-330 <RA2>
A:Cross-references: GB:S39291; NID:9250908; PIDN:AA82446.1; PID:9250910
A:Experimental source: B4; virulent strain
A:Note: sequence extracted from NCBI backbone (NCBIN:108105, NCBIP:108107)
C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase; F
F:1-69/Product: coat protein 1A #status predicted <C1A>
F:1-70-330/Product: coat protein 1B #status predicted <C1B>
F:331-568/Product: coat protein 1C #status predicted <C1C>
F:569-852/Product: coat protein 1D #status predicted <C1D>
F:853-999/Product: core protein 2A #status predicted <C2A>
F:1000-1098/Product: core protein 2B #status predicted <C2B>
F:1099-1427/Product: core protein 2C #status predicted <C2C>
F:1428-1516/Product: protein 3A #status predicted <P3A>
F:1517-1558/Product: genome-linked protein VPg #status predicted <VPG>
F:1553-1721/Product: proteinase #status predicted <PPS>
F:1722-2183/Product: RNA-directed RNA polymerase #status predicted <RPS>
F:1519/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2183;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1423 EALFOG 1428

RESULT 7
GNNYB3
genome polypeptide - coxsackievirus B3 (strain Nancy)
N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; cor
EC 3.4.-.-; RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: coxsackievirus B3
A:Note: host Homo sapiens (man)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 05-Jun-1998
A:Accession: A26354; B00725
R.Lindberg, A.M.; Stalhandske, P.O.K.; Petersson, U.
Virology 156, 50-63, 1987
A:Title: Genome of coxsackievirus B3
A:Reference number: A26354; MUID:87122156; PMID:3027968
A:Accession: A26354
A:Molecule type: genomic RNA
A:Residues: 1-2185 <LIN>
A:Cross-references: GB:K02709
R.Stalhandske, P.O.K.; Lindberg, M.; Petersson, U.
J. Virol. 51, 742-746, 1984
A:Title: Replicase gene of coxsackievirus B3.
A:Reference number: A00725; MUID:84292451; PMID:6088796
A:Accession: B00725
A:Molecule type: genomic RNA
A:Residues: 1724-1733; 'D', 1735-2185 <STA>
A:Cross-references: GB:M16572

C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotid
F:1-69/Product: coat protein VP4 #status predicted <VP4>
F:1-70-332/Product: coat protein VP2 #status predicted <VP2>
F:333-570/Product: coat protein VP3 #status predicted <VP3>
F:571-880/Product: coat protein VP1 #status predicted <VP1>
F:881-1001/Product: core protein P2-A #status predicted <P2A>
F:1002-1100/Product: core protein P2-B #status predicted <P2B>
F:1101-1429/Product: core protein P2-C #status predicted <P2C>
F:1430-1518/Product: core protein P3-A #status predicted <P3A>
F:1519-1540/Product: genome-linked protein VPg #status predicted <VPG>
F:1541-1723/Product: proteinase #status predicted <PPS>
F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <RPS>
F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1425 EALFOG 1430

RESULT 9
JQ2021
genome polypeptide - coxsackievirus B5 (strain 1954/UK/85)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; cor
polymerase (EC 2.7.7.48)
C:Species: coxsackievirus B5
C:Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 16-Jul-1999
A:Accession: JQ2021
R.Zhang, G.; Wilsden, G.; Knowles, N.J.; McCauley, J.W.
A:Title: Complete nucleotide sequence of infectious coxsackievirus B3 CDNA: two int
A:Reference number: A34664; MUID:90204667; PMID:2157045
A:Accession: A34664
A:Molecule type: genomic RNA
A:Residues: 1-2185 <KLU>
A:Cross-references: GB:M33854; NID:9323419; PIDN:AAA42931.1; PID:9323420
C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransfera
F:1-69/Product: coat protein 1A #status predicted <VP4>
F:1-70-332/Product: coat protein 1B #status predicted <VP2>
F:333-570/Product: coat protein 1C #status predicted <VP3>
F:571-880/Product: coat protein 1D #status predicted <VP1>
F:881-1001/Product: core protein 2A #status predicted <P2A>
F:1002-1100/Product: core protein 2B #status predicted <P2B>
F:1101-1429/Product: core protein 2C #status predicted <P2C>
F:1430-1518/Product: protein 3A #status predicted <P3A>
F:1519-1540/Product: genome-linked protein VPg #status predicted <VPG>
F:1541-1723/Product: proteinase #status predicted <PPS>
F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <RPS>
F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

J. Gen. Virol. 74, 845-853, 1993
 A>Title: Complete nucleotide sequence of a coxsackie B5 virus and its relationship to sw
 A:Reference number: J02021; MUID:93260398; PMID:8388019
 A:Accession: J02021
 A:Molecule type: genomic RNA
 A:Residues: 1-2185 <ZHA>
 A:Cross-References: GB:X67706; NID:959045; PIDN:CAA47944.1; PID:959046
 C:Superfamily: poliovirus genome polyprotein
 C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase; P
 F:1-69/Product: coat protein 1A #status predicted <C1A>
 F:70-330/Product: coat protein 1B #status predicted <C1B>
 F:331-566/Product: coat protein 1C #status predicted <C1C>
 F:566-831/Product: coat protein 1D #status predicted <C1D>
 F:852-1001/Product: core protein 2A #status predicted <C2A>
 F:1002-1100/Product: core protein 2B #status predicted <C2B>
 F:1101-1429/Product: core protein 2C #status predicted <C2C>
 F:1430-1518/Product: core protein 3A #status predicted <P3A>
 F:1519-1540/Product: genome-linked protein VPg #status predicted <VPg>
 F:1541-1723/Product: proteinase #status predicted <P3>
 F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <RPS>
 F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 10
 GNNYSV
 genome polyprotein - swine vesicular disease virus (strain UKG/27/72)
 N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core
 A-directed RNA polymerase (EC 2.7.7.48)
 C:Species: swine vesicular disease virus
 C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
 C:Accession: S11670; A30061
 R:Seechurn, P.; Knowles, N.J.; McCauley, J.W.
 Virus Res. 16, 255-274, 1990
 A>Title: The complete nucleotide sequence of a pathogenic swine vesicular disease virus.
 A:Reference number: S11670; MUID:90364770; PMID:2168111
 A:Molecule type: genomic RNA
 A:Residues: 1-2185 <SEE>
 A:Cross-References: EMBL:X54521; NID:961167; PIDN:CAA38377.1; PID:961168
 A:Note: submitted to the Protein Sequence Database, April 1989
 C:Superfamily: poliovirus genome polyprotein
 C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotidyltra
 F:1-69/Product: coat protein VP4 #status predicted <VP4>
 F:70-330/Product: coat protein VP2 #status predicted <VP2>
 F:331-566/Product: coat protein VP3 #status predicted <VP3>
 F:566-851/Product: coat protein VP1 #status predicted <VP1>
 F:852-1001/Product: core protein P2-3b #status predicted <P2A>
 F:1002-1100/Product: core protein P2-3b #status predicted <P2A>
 F:1101-1429/Product: core protein P2-3b #status predicted <P2B>
 F:1430-1518/Product: core protein P3-1b #status predicted <P3A>
 F:1519-1540/Product: genome-linked protein VPg #status predicted <P3B>
 F:1541-1723/Product: proteinase 3C #status predicted <P3C>
 F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <P3D>
 F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 11

GNNYSV
 genome polyprotein - swine vesicular disease virus (strain H/3/76)
 N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; c
 A-directed RNA polymerase (EC 2.7.7.48)
 C:Species: swine vesicular disease virus
 C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jun-2000
 C:Accession: A31331
 R:Inoue, T.; Suzuki, T.; Sekiguchi, K.
 J. Gen. Virol. 70, 919-934, 1989
 A>Title: The complete nucleotide sequence of swine vesicular disease virus.
 A:Reference number: A31331; MUID:89279274; PMID:2543767
 A:Accession: A31331
 A:Molecule type: genomic RNA
 A:Residues: 1-2185 <INO>
 A:Cross-References: GB:D00435; NID:91228947; PIDN:BA00337.1; PID:9222630
 C:Superfamily: poliovirus genome polyprotein
 C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotidyl
 F:1-69/Product: coat protein VP4 #status predicted <VP4>
 F:70-330/Product: coat protein VP2 #status predicted <VP2>
 F:331-566/Product: coat protein VP3 #status predicted <VP3>
 F:566-851/Product: coat protein VP1 #status predicted <VP1>
 F:852-1001/Product: core protein P2-3b #status predicted <P2A>
 F:1002-1100/Product: core protein P2-3b #status predicted <P2B>
 F:1101-1429/Product: core protein P2-X #status predicted <P2C>
 F:1430-1518/Product: protein P3-1b #status predicted <P3A>
 F:1519-1540/Product: genome-linked protein VPg #status predicted <P3B>
 F:1541-1723/Product: proteinase 3C #status predicted <P3C>
 F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <P3D>
 F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 12
 S44194
 polyprotein - echovirus 12
 C:Species: echovirus 12
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
 C:Accession: S44194
 R:Kraus, W.; Nielsen-Salzs, B.E.
 submitted to the EMBL Data Library, February 1994
 A:Reference number: S44194
 A:Accession: S44194
 A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-2193 <KRA>
 A:Cross-References: EMBL:X77708; NID:92058304; PIDN:CAA54783.1; PID:9474421
 C:Superfamily: poliovirus genome polyprotein
 C:Keywords: genome-linked protein; phosphoprotein; polyprotein
 F:1529/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 2; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1433 EALFOG 1438

RESULT 13
 S44251
 polyprotein - echovirus 12
 C:Species: echovirus 12
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Jun-1998
 C:Accession: S44251
 R:Zimmermann, H.; Kraus, W.; Nielsen-Salzs, B.E.

Submitted to the EMBL Data Library, February 1994

A:Reference number: S44251

A:Accession: S44251

A>Status: Preliminary

A:Molecule type: genomic RNA

A:Residues: 1-2193 <ZIM>

A:Cross-references: EMBL:X79047

C:Superfamily: poliovirus genome polyprotein

C:Keywords: genome-linked protein; phosphoprotein; polyprotein

F:1529/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match

Best Local Similarity 100.0%; Score 30; DB 2; Length 2193;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6

Db 1433 EALFOG 1438

RESULT 14

S52919

polyprotein (1A, 1B, 1C, 1D, 2A, 2B, 2C, 3A, 3B, 3C, 3D) - echovirus 9

C:Species: echovirus 9

C:Date: 06-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 05-Jun-1998

C:Accession: S52919

R:Zimmermann, H.; Eggers, H.J.; Nelsen-Salz, B.

submitted to the EMBL Data Library, February 1995

A:Description: Complete nucleotide sequence and construction of an infectious clone of E

A:Reference number: S52919

A:Accession: S52919

A>Status: Preliminary

A:Molecule type: genomic RNA

A:Residues: 1-2193 <ZIM>

A:Cross-references: EMBL:X84981

C:Superfamily: poliovirus genome polyprotein

C:Keywords: genome-linked protein; phosphoprotein; polyprotein

F:1529/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match

Best Local Similarity 100.0%; Score 30; DB 2; Length 2193;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6

RESULT 15

GNNY7

genome polyprotein - human enterovirus 70 (strain J670/71)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro

NA-directed RNA polymerase (EC 2.7.7.48)

C:Species: human enterovirus 70

C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 05-Jun-1998

C:Accession: A36253

R:Ryan, M.D.; Jenkins, O.; Hughes, P.J.; Brown, A.; Knowles, N.J.; Booth, D.; Minor, P.D.

J. Gen. Virol. 71, 2291-2299, 1990

A:Title: The complete nucleotide sequence of enterovirus type 70: relationships with oth

A:Reference number: A36253; M01D:91037960; PMID:2172447

A:Accession: A36253

A:Molecule type: genomic RNA

A:Residues: 1-2194 <RYA>

A:Cross-references: GB:D00820

C:Superfamily: poliovirus genome polyprotein

C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotidyltra

F:1-65/Product: coat protein 1A #status predicted <VP4>

F:70-319/Product: coat protein 1B #status predicted <VP2>

F:320-561/Product: coat protein 1C #status predicted <VP3>

F:562-871/Product: coat protein 1D #status predicted <VP1>

F:872-1014/Product: core protein 2A #status predicted <P2A>

F:1015-1113/Product: core protein 2B #status predicted <P2B>

F:1114-1443/Product: core protein 2C #status predicted <P2C>

F:1444-1532/Product: protein 3A #status predicted <P3A>

F:1533-1554/Product: genome-linked protein VPg #status predicted <VPg>

F:1555-1737/Product: proteinase 3C #status predicted <P3C>

F:1738-2194/Product: RNA-directed RNA polymerase #status predicted <RRP>

F:1535/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match

Best Local Similarity 100.0%; Score 30; DB 1; Length 2194;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6

Db 1439 EALFOG 1444

Search completed: May 20, 2003, 11:39:25
Job time: 1.2315 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 0.141625 Seconds

(without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30

Sequence: 1 EALFOG 6

Scoring table: BLOSUM62

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	1081	1	PDR6_YEAST
2	30	100.0	2175	1	POLG_BOVEY
3	30	100.0	2182	1	POLG_CXB1J
4	30	100.0	2183	1	POLG_CXB4E
5	30	100.0	2183	1	POLG_CXB4J
6	30	100.0	2184	1	POLG_EC01F
7	30	100.0	2185	1	POLG_CXB3N
8	30	100.0	2185	1	POLG_CXB3W
9	30	100.0	2185	1	POLG_CXB5P
10	30	100.0	2185	1	POLG_SVDVH
11	30	100.0	2191	1	POLG_EC06C
12	30	100.0	2193	1	POLG_CXA16
13	30	100.0	2193	1	POLG_EC09H
14	30	100.0	2193	1	POLG_EC12T
15	30	100.0	2193	1	POLG_HE71B
16	30	100.0	2193	1	POLG_HE71M
17	30	100.0	2194	1	POLG_EC30B
18	30	100.0	2194	1	POLG_EC701
19	30	100.0	2195	1	POLG_EC05N
20	30	100.0	2196	1	POLG_EC09B
21	30	100.0	2201	1	POLG_CXA9
22	30	100.0	2203	1	POLG_EC09B
23	30	100.0	2206	1	POLG_CXA21
24	30	100.0	2206	1	POLG_POL1M
25	30	100.0	2206	1	POLG_POL1M
26	30	100.0	2206	1	POLG_POL1M
27	30	100.0	2207	1	POLG_POL1M
28	30	100.0	2207	1	POLG_POL1M
29	30	100.0	2209	1	POLG_POL1M
30	30	100.0	2214	1	POLG_CXA24
31	30	100.0	2214	1	POLG_HRV89
32	30	93.3	2164	1	POLG_CXA24
33	27	90.0	272	1	COX3_PYL1I

34	27	90.0	323	1	TTD2_HUMAN	P09758	homo sapien
35	27	90.0	365	1	FXH1_HUMAN	O75593	homo sapien
36	27	90.0	395	1	SYW_A00AE	O67115	aquifex aeo
37	27	90.0	443	1	IF5_PHAVU	P48724	phaseolus v
38	27	90.0	452	1	HIS7_PHYPR	P28624	phytophthor
39	27	90.0	614	1	SPAT_BACSU	P33116	baecillus su
40	27	90.0	707	1	HLV2_ECOLI	P10089	escherichia
41	27	90.0	707	1	LYB8_ECOLI	P08716	escherichia
42	27	90.0	1337	1	YDM5_SCHPO	P87136	schizosacch
43	26	86.7	167	1	FTN_CAMJE	O46106	campylobact
44	26	86.7	196	1	HIS7_HA1N1	O9h13	halobacteri
45	26	86.7	201	1	CYC4_PSEAE	P00106	pseudomonas

ALIGNMENTS

RESULT 1

ID	PDR6_YEAST	STANDARD:	PRT:	1081 AA.
AC	P32767;			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Pleiotropic drug resistance regulatory protein 6.			
GN	PDR6 OR YGL016W.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=IL125-2B;			
RA	MEDLINE=91353083; PubMed=1882553;			
RX	Chen W., Balzi E., Capleaux E., Choder M., Goffeau A.;			
RT	"The DNA sequencing of the 17 kb HindIII fragment spanning the LEU1			
RT	and ATR1 loci on chromosome VII from Saccharomyces cerevisiae reveals			
RT	the pDR6 gene, a new member of the genetic network controlling			
RT	pleiotropic drug resistance.";			
RL	Yeast 7:287-299(1991)..			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Hebling U., Hofmann B., Dellus H.;			
RL	Submitted (MAY 1996) to the EMBL/Genbank/DBJ databases.			
CC	-1- FUNCTION: PROBABLE POSITIVE REGULATOR OF PROTEINS INVOLVED IN			
CC	PERMEABILITY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: S58126; AAD13801.1; -;			
DR	EMBL: S57895; AAB19613.1; -;			
DR	EMBL: Z72538; CAA96716.1; -;			
DR	PIR: S15040; S15040.			
DR	SGD: S0002984; KAP122.			
DR	SEQUENCE 1081 AA; 123530 MW; EEBG5ADD4E16D9D3 CRC64;			
QY	Query Match	100.0%;	Score 30;	DB 1; Length 1081;
QY	Best Local Similarity	100.0%;	Pred. No. 26;	
QY	Matches 6;	Conservative 0;	Mismatches 0;	Indels 0;
QY	1 EALFOG 6			
QY				
QY	631 EALFOG 636			
DB				
DB	POLG_BOVEY			

RESULT 2

1D POLG_BOVEV STANDARD; PRT; 2175 AA.
 AC P12915;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C, P3A; Genome-linked protein VPg; Picornain 3C
 DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
 DE (EC 2.7.7.48)]
 OS Bovine enterovirus (strain VG-5-27) (BEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 ON NCBI_TaxID=12065;
 RX MEDLINE=88117392; PubMed=2828511;
 RA Earle J.A.P., Skuce R.A., Fleming C.S., Hoey E.M., Martin S.J.;
 RT "The complete nucleotide sequence of a bovine enterovirus.";
 RL J. Gen. Virol. 69:253-263(1988).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 1-840.
 RX MEDLINE=95292108; PubMed=773791;
 RA Smyth M., Tate J., Hoey E.M., Lyons C., Martin S.J., Stuart D.;
 RT "Implications for viral uncoating from the structure of bovine enterovirus.";
 RL Nat. Struct. Biol. 2:224-231(1995).
 CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 CC O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the
 CC poliovirus polypeptide. In other picornavirus reactions Gln may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 CC EMBL: D00214; BAA24003.1; ALT_SEQ.
 DR PIR: A29824; GNNYBE.
 DR PDB: 1BEV; 16-SEP-98.
 DR MEROPS: C03 UPA; -;
 DR MEROPS: C03 UPA; -;
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR000318; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rnv.
 DR Pfam: PF00073; Inv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KM Polyprotein: Coat protein; Core protein; Transferase;
 KM RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
 KM 3D-structure.
 FT CHAIN 2 69 COAT PROTEIN VP4 (P1A).

FT CHAIN 70 317 COAT PROTEIN VP2 (P1B).
 FT CHAIN 318 559 COAT PROTEIN VP3 (P1C).
 FT CHAIN 560 840 COAT PROTEIN VP1 (P1D).
 FT CHAIN 841 990 CORE PROTEIN P2A.
 FT CHAIN 991 1088 CORE PROTEIN P2B.
 FT CHAIN 1090 1419 CORE PROTEIN P2C.
 FT CHAIN 1420 1508 CORE PROTEIN P3A.
 FT CHAIN 1509 1531 GENOME-LINKED PROTEIN VPg (P3B).
 FT CHAIN 1532 1714 PICORNAIN 3C.
 FT CHAIN 1715 2175 RNA-DEPENDENT RNA POLYMERASE P3D.
 FT LIPID 2 2 MYRISTATE.
 FT ACT_SITE 1678 1678 PROTEASE (POTENTIAL).
 FT ACT_SITE 1692 1692 PROTEASE (POTENTIAL).
 SQ SEQUENCE 2175 AA; 242502 MM; 44FCADB8704E48FD CRC64;
 QY 1 EALFOG 6 100.0%; Score 30; DB 1; Length 2175;
 Db 1415 EALFOG 1420 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 3
 ID POLG_CXB1J STANDARD; PRT; 2182 AA.
 AC P08291;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
 DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 OS Coxsackievirus B1 (strain Japan).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 ON NCBI_TaxID=103902;
 RX MEDLINE=87122157; PubMed=3027969;
 RA Itzuka N., Kuge S., Nomoto A.;
 RT "Complete nucleotide sequence of the genome of coxsackievirus B1.";
 RL Virology 156:64-73(1987).
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN O/G SITES IN THE POLYPEPTIDE. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the
 CC poliovirus polypeptide. In other picornavirus reactions Gln may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the
 CC picornavirus polypeptide. In other picornavirus reactions Gln may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS
 CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16560; AAC00531.1; -
DR PIR: A26353; GNNVBL.
DR HSSP: P21404; IDAM.
DR MEROPS: C03.011; -
DR MEROPS: C03.022; -
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rhy.
DR Pfam: PF00073; Rhy; 3.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
DR Polyprotein; Coat protein; Core protein; Transferase; Myristate;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2 69
FT CHAIN 70 332
FT CHAIN 333 570
FT CHAIN 571 848
FT CHAIN 849 998
FT CHAIN 999 1097
FT CHAIN 1098 1426
FT CHAIN 1427 1515
FT CHAIN 1516 1537
FT CHAIN 1538 1720
FT CHAIN 1721 2182
FT LIPID 2 2
FT ACT_SITE 1684 1684
FT ACT_SITE 1698 1698
SQ SEQUENCE 2182 AA; 243945 MW; 387B939127585B1 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2182;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALF0G 6
Db 1422 EALF0G 1427

RESULT 4
ID POLG_CXB4E STANDARD; PRT; 2183 AA.
AC 086887;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
(P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
(P2A); Coat protein VP2 (P2B); Core protein P2C; Core protein
(P3A); Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28)
DE (EC 3.4.22.28) (P2A); Core protein P2B; Core protein P2C; Core protein
(P3A); Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28)
OS Coxsackievirus B4 (strain E2).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=103905;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95205102; PubMed=7897366;
RA Kang Y., Chatterjee N.K., Nodwell M.J., Yoon J.W.;
*Complete nucleotide sequence of a strain of coxsackie B4 virus of

RT human origin that induces diabetes in mice and its comparison with
RT nondiabetogenic coxsackie B4 JEV strain.";
RL J. Med. Virol. 44:353-361(1994).
CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
CC AT CERTAIN O/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
CC PROTEASES.
CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Glu-1-Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Glu may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the
CC picornavirus polyprotein. In other picornavirus reactions Glu may
CC be substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N').
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/VP2A IS
CC CATALYZED BY P2A. ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S76772; AB33885.1; -
DR HSSP: P21404; IDAM.
DR InterPro: IPR004004; Calicl_pol_hel.
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rhy.
DR Pfam: PF00073; Rhy; 3.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR PRINTS: PR00918; CALICIVIRUSNS.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
DR Polyprotein; Coat protein; Core protein; Transferase; Myristate;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2 69
FT CHAIN 70 330
FT CHAIN 331 568
FT CHAIN 569 849
FT CHAIN 850 999
FT CHAIN 1000 1098
FT CHAIN 1099 1427
FT CHAIN 1428 1516
FT CHAIN 1517 1538
FT CHAIN 1539 1721
FT CHAIN 1722 2183
FT LIPID 2 2
FT ACT_SITE 1685 1685
FT ACT_SITE 1699 1699
SQ SEQUENCE 2183 AA; 244649 MW; BF44781CF6981D39 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2183;

Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 1423 EALFOG 1428

RESULT 5
POLG_CXB4J STANDARD; PRT: 2183 AA.
ID POLG_CXB4J
AC P08292;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome: polypeptide (Contains: Coat protein VP1 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP4 (P1D); Picornain 2A (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)).
DE Coxsackievirus B4 (strain JVB / Benschooten / New York/51).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Enterovirus.
OC NCBI_TaxID=103906;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87253111; PubMed=3037008;
RA Jenkins O., Booth J.D., Minor P.D., Almond J.W.;
RT "The complete nucleotide sequence of coxsackievirus B4 and its comparison to other members of the Picornaviridae.";
RL J. Gen. Virol. 68:1835-1848 (1987).
RN [2]
RP REVISIONS.
RA Jenkins O.;
RL Submitted (AUG-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE AT CERTAIN Q/G SITES IN THE POLYPEPTIDE. THEY ARE CYSTEINE PROTEASES.
CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the poliovirus polypeptide. In other picornavirus reactions Gln may be substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the picornavirus polypeptide. In other picornavirus reactions Gln may be substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + [RNA](N).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS. CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib.ch).
CC -----
DR EMBL: X05690; CAA29172.1; -;
DR PIR: A27170; GNNYB4.
DR HSSP: P21404; 1D4M.
DR MEROPS: C03.011; -;
DR MEROPS: C03.022; -;
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.

DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001576; Rnv.
DR Pfam: PF00073; Rnv; 3.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
KM Polypeptide: Coat protein; Core protein; Transferase; Myristate;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2
FT CHAIN 69
FT CHAIN 70 330
FT CHAIN 331 568
FT CHAIN 569 849
FT CHAIN 850 999
FT CHAIN 1000 1098
FT CHAIN 1099 1427
FT CHAIN 1428 1516
FT CHAIN 1517 1538
FT CHAIN 1539 1721
FT CHAIN 1722 2183
FT CHAIN 2183
FT LIPID 2
FT ACT_SITE 1685 1685
FT ACT_SITE 1699 1699
FT ACT_SITE 1699 1699
SQ SEQUENCE 2183 AA; 244012 MW; 244012 MW; DE069DE3AE91AF0 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 1; Length 2183;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 1423 EALFOG 1428

RESULT 6
POLG_EC01F STANDARD; PRT: 2184 AA.
ID POLG_EC01F
AC 091734; 066795; 09YID6;
DT 15-JUL-1998 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Genome: polypeptide (Contains: Coat protein VP4 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)).
DE Echovirus 1 (strain Farok / ATCC VR-1038).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Enterovirus.
OC NCBI_TaxID=103908;
RN [1]
RP SEQUENCE FROM N.A.
RA Bergelson J.M.;
RT "Receptor interactions, infectious cDNA, and nucleotide sequences of echovirus 1/8.";
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 69-330 FROM N.A.
RX MEDLINE=96203959; PubMed=8627260;
RA Hutunen P., Santti J., Pulli T., Hyypia T.;
RT "The major echovirus group is genetically coherent and related to coxsackie B viruses.";
RL J. Gen. Virol. 77:715-725 (1996).
RN [3]
RP SEQUENCE OF 570-853 FROM N.A.
RX MEDLINE=99138973; PubMed=9971773;

RA Oberste M.S., Maher K., Kilpatrick D.R., Pallansch M.A.:
 RT "Molecular evolution of the human enteroviruses: correlation of
 RT serotype with vp1 sequence and application to picornavirus
 RT classification." J. Virol. 73:1941-1948(1999).
 RN (4)
 RP X-RAY CRYSTALLOGRAPHY (3.55 ANGSTROMS) OF 2-850.
 RX MEDLINE-99192672; PubMed-10089503;
 RA Filman D.J., Wien M.W., Cunningham J.A., Bergelson J.M., Hogle J.M.:
 RT "Structure determination of echovirus 1." J. Virol. 73:1941-1948(1999).
 RL Acta Crystallogr. D 54:1261-1272(1998).
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- FUNCTION: It is thought that the P2C protein attaches to vesicular
 CC membranes and is associated with viral RNA synthesis.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-gly bond in the
 CC picornavirus polyprotein. In other picornavirus reactions Glu may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC. VP1/P2A IS
 CC CATALYZED BY P2A. ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF029859; AAC63944.2;
 CC EMBL: X89531; CA61710.1;
 CC EMBL: AF081314; AAD17718.1;
 CC PDB: 1EVI; 27-JAN-99.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR004004; Calic_Pol_hel.
 DR InterPro: IPR000199; Cys-Protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; RNv.
 DR Pfam: PF00073; thv; 4.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR PRINTS: PR00918; CALICITRUSNS.
 DR PRODOM: PD001125; Cys-Protease-3C; 1.
 DR PRODOM: PD001274; Pico_P2B; 1.
 DR PRODOM: PD001306; Pico_P2A; 1.
 DR SMART: SM00382; AAA; 1.
 KM Polyprotein; Coat protein; Core protein; Transferase; Myristate;
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.
 FT CHAIN 2
 FT CHAIN 70
 FT CHAIN 330
 FT CHAIN 569
 FT CHAIN 570
 FT CHAIN 850
 FT CHAIN 851
 FT CHAIN 1000
 FT CHAIN 1001
 FT CHAIN 1009
 FT CHAIN 1100
 FT CHAIN 1428
 FT CHAIN 1429
 FT CHAIN 1517
 FT CHAIN 1518
 FT CHAIN 1539
 FT CHAIN 1540
 FT CHAIN 1722
 FT CHAIN 1723
 FT CHAIN 2184
 FT CHAIN 2
 FT CHAIN 2
 LIPID 2
 LIPID 2
 MYRISTATE.
 RNA-DIRECTED RNA POLYMERASE.
 RNA-DIRECTED RNA POLYMERASE.
 MYRISTATE.

FT ACT_SITE 1686 1686 PROTEASE 3C (POTENTIAL).
 FT ACT_SITE 1700 1700 PROTEASE 3C (POTENTIAL).
 FT CONFLICT 69 69 N -> M (IN REF. 2).
 FT CONFLICT 238 238 A -> R (IN REF. 2).
 FT CONFLICT 611 611 A -> V (IN REF. 3).
 FT CONFLICT 645 645 Y -> F (IN REF. 3).
 FT CONFLICT 724 724 V -> I (IN REF. 3).
 FT CONFLICT 774 774 A -> T (IN REF. 3).
 SO SEQUENCE 2184 AA; 244003 MM; CBB937D0AC18EA9 CRC64;
 Query Match 100.0%; Score 30; DB 1; Length 2184;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EALFQG 6
 DB 1424 EALFQG 1429
 RESULT 7
 ID POLG_CXB3N STANDARD; PRT: 2185 AA.
 AC P03313; Q66322; Q66323; Q66324; Q66325; Q66326; Q66327; Q66328;
 AC Q83744;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
 DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 OS Coxsackievirus B3 (strain Nancy).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxID=103903;
 RN [1]
 RN SEQUENCE FROM N.A.
 RN MEDLINE-90204667; PubMed-2157045;
 RA Klump W.M., Bergmann I., Mueller B.C., Ameis D., Kandolf R.:
 RT "Complete nucleotide sequence of infectious Coxsackievirus B3 CDNA:
 RT two initial 5' uridine residues are regained during plus-strand RNA
 RT synthesis." J. Virol. 64:1573-1583(1990).
 RL J. Virol. 64:1573-1583(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RN MEDLINE-8712156; PubMed-3027968;
 RA Lindberg A.M., Staalhandske P.O.K., Pettersson U.:
 RT "Genome of coxsackievirus B3." Virology 156:50-63(1987).
 RL Virology 156:50-63(1987).
 RN [3]
 RN SEQUENCE OF 1724-2185 FROM N.A.
 RN MEDLINE-84292451; PubMed-6088796;
 RA Staalhandske P.O.K., Lindberg A.M., Pettersson U.:
 RT "Replicase gene of coxsackievirus B3." J. Virol. 51:742-746(1984).
 RL J. Virol. 51:742-746(1984).
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-851.
 RA Muckelbauer J.K., Kremer M., Minor I., Tong L., Zlotnick A.,
 RA Johnson J.E., Rossmann M.G.:
 RT "Structure determination of coxsackievirus B3 to 3.5-A resolution." J.
 RT Acta Crystallogr. D 51:871-887(1995).
 RL Acta Crystallogr. D 51:871-887(1995).
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-gly bond in the
 CC picornavirus polyprotein. In other picornavirus reactions Glu may
 CC be substituted for Gln, and Ser or Thr for Gly.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PMV: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC: VP1/VP2 IS
 CC CATALYZED BY P2A. ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M33854; AAA42931.1; -;
 DR EMBL: K02709; AAA42932.1; -;
 DR EMBL: M16572; AAA74400.1; -;
 DR PIR: A26354; GNNYB3.
 DR PIR: A34654; GNNYBT.
 DR PDB: 1COV; 06-MAR-96.
 DR MEROPS: C03.011; -;
 DR MEROPS: C03.022; -;
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR000138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; RNv.
 DR Pfam: PF00073; RNv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KM Polyprotein; Coat protein; Core protein; Transferase; Myristate;
 KM RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.
 FT CHAIN 2
 FT CHAIN 69
 FT CHAIN 70 332
 FT CHAIN 333 570
 FT CHAIN 571 851
 FT CHAIN 852 1001
 FT CHAIN 1002 1100
 FT CHAIN 1101 1429
 FT CHAIN 1430 1518
 FT CHAIN 1519 1540
 FT CHAIN 1541 1723
 FT CHAIN 1724 2185
 FT LIPID 2
 FT ACT_SITE 1687
 FT ACT_SITE 1701 1701
 FT CONFLICT 16 16
 FT CONFLICT 17 17
 FT CONFLICT 469 469
 FT CONFLICT 487 487
 FT CONFLICT 510 510
 FT CONFLICT 516 516
 FT CONFLICT 566 566
 FT CONFLICT 593 593
 FT CONFLICT 650 650
 FT CONFLICT 854 865
 FT CONFLICT 873 1097
 FT CONFLICT 1097 1280
 FT CONFLICT 1280 1280

FT CONFLICT 1437 1437 I -> F (IN REF. 2).
 FT CONFLICT 1503 1503 V -> M (IN REF. 2).
 FT CONFLICT 1616 1616 K -> E (IN REF. 2).
 FT CONFLICT 1624 1624 R -> G (IN REF. 2).
 FT CONFLICT 1627 1627 R -> G (IN REF. 2).
 FT CONFLICT 1630 1630 L -> V (IN REF. 2).
 FT CONFLICT 1718 1718 Y -> N (IN REF. 2).
 FT CONFLICT 1734 1734 D -> V (IN REF. 2).
 FT CONFLICT 1758 1758 E -> V (IN REF. 2 AND 3).
 FT CONFLICT 1824 1824 V -> R (IN REF. 2 AND 3).
 FT CONFLICT 1867 1867 C -> R (IN REF. 2 AND 3).
 FT CONFLICT 1880 1880 Y -> H (IN REF. 2 AND 3).
 FT CONFLICT 2001 2001 D -> N (IN REF. 2 AND 3).
 FT CONFLICT 2095 2095 A -> V (IN REF. 2 AND 3).
 FT CONFLICT 2115 2115 V -> A (IN REF. 2 AND 3).
 FT CONFLICT 2175 2175 S -> T (IN REF. 2 AND 3).
 FT CONFLICT 2178 2178 R -> G (IN REF. 2 AND 3).
 SQ SEQUENCE 2185 AA; 243451 MW; 1BSCE3DA7338FF CRC64;
 Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 EALFQG 6
 Db 1425 EALFQG 1430
 ID: PICO_C3B3M STANDARD: PRT: 2185 AA.
 AC Q66282; 15-0UL-1999 (Rel. 38, Created)
 DT 15-0UL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
 DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 OS Coxsackievirus B3 (strain Woodruff).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxID=103904;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Knowlton K.U., Jeon E.S., Berkley R.W., Messely R., Huber S.;
 RT "A mutation in the puff region of VP2 attenuates the myocarditic
 RT phenotype of an infectious cDNA of the Woodruff virus";
 RT Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Glu-1-gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-gly bond in the
 CC picornavirus polyprotein. In other picornavirus reactions Glu may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PMV: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC: VP1/VP2 IS
 CC CATALYZED BY P2A. ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: U57056; AAB02228.1; -
 DR HSSP: Q83744; LCOV.
 DR MEROPS: C03.011; -
 DR MEROPS: C03.022; -
 DR InterPro: IPR004004; Calic1.pol.hel.
 DR InterPro: IPR000199; Cys-protease-3c.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rnv.
 DR Pfam: PF00073; Rnv; 3.
 DR Pfam: PF00548; Cys-protease-3c; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR PRINTS: PR00918; CALICIVIRUSNS.
 DR ProDom: PD001125; Cys-protease-3c; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease;
 FT CHAIN 2 69 COAT PROTEIN VP4.
 FT CHAIN 70 332 COAT PROTEIN VP2.
 FT CHAIN 333 570 COAT PROTEIN VP3.
 FT CHAIN 571 851 COAT PROTEIN VP1.
 FT CHAIN 852 1001 PICORNAIN 2A.
 FT CHAIN 1002 1100 CORE PROTEIN P2B.
 FT CHAIN 1101 1429 CORE PROTEIN P2C.
 FT CHAIN 1430 1518 CORE PROTEIN P3A.
 FT CHAIN 1519 1540 GENOME-LINKED PROTEIN VP6.
 FT CHAIN 1541 1723 PICORNAIN 3C.
 FT CHAIN 1724 2185 RNA-DIRECTED RNA POLYMERASE.
 FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
 FT ACT_SITE 1687 1687 PROTEASE 3C (POTENTIAL).
 FT ACT_SITE 1701 1701 PROTEASE 3C (POTENTIAL).
 SQ SEQUENCE 2185 AA; 243680 MM; FD93A677904252FA CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EALFQG 6
 Db 1425 EALFQG 1430

RESULT 9
 POLG_CXB5P
 ID POLG_CXB5P STANDARD; PRT; 2185 AA.
 AC Q03053;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
 DE (P2C 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 OS Coxsackievirus B5 (strain Peterborough / 1954/UK/85).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxID=103907;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93260398; PubMed-8388019;
 RA Zhang G., Wilsden G., Knowles N.J., McCauley J.W.;
 RT "Complete nucleotide sequence of a coxsackie B5 virus and its
 RT relationship to swine vesicular disease virus.";
 RL J. Gen. Virol. 74:845-853(1993).
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN O/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Gln may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-gly bond in the
 CC picornavirus polyprotein. In other picornavirus reactions Gln may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS
 CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X67706; CA47944.1; -
 DR PIR: J02021; J02021.
 DR HSSP: P21404; ID4M.
 DR MEROPS: C03.011; -
 DR MEROPS: C03.022; -
 DR InterPro: IPR000199; Cys-protease-3c.
 DR InterPro: IPR00138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rnv.
 DR Pfam: PF00073; Rnv; 3.
 DR Pfam: PF00548; Cys-protease-3c; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3c; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease;
 FT CHAIN 2 69 COAT PROTEIN VP4.
 FT CHAIN 70 330 COAT PROTEIN VP2.
 FT CHAIN 331 568 COAT PROTEIN VP3.
 FT CHAIN 569 851 COAT PROTEIN VP1.
 FT CHAIN 852 1001 PICORNAIN 2A.
 FT CHAIN 1002 1100 CORE PROTEIN P2B.
 FT CHAIN 1101 1429 CORE PROTEIN P2C.
 FT CHAIN 1430 1518 CORE PROTEIN P3A.
 FT CHAIN 1519 1540 GENOME-LINKED PROTEIN VP6.
 FT CHAIN 1541 1723 PICORNAIN 3C.
 FT CHAIN 1724 2185 RNA-DIRECTED RNA POLYMERASE.
 FT LIPID 2 2 MYRISTATE (BY SIMILARITY).

FT ACT_SITE 1687 1687 PROTEASE 3C (POTENTIAL).
 FT ACT_SITE 1701 1701 PROTEASE 3C (POTENTIAL).
 SO SEQUENCE 2185 AA; 243298 MW; 3F9EE29F90D59C6F CRC64;

Query Match
 Best Local Similarity 100.0%; Score 30; DB 1; Length 2185;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 10
 POLG_SVDVH STANDARD; PRT; 2185 AA.

AC P1604;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polypeptide (Contains: Coat proteins VP1 TO VP4; Core proteins
 DE P2A TO P2C; P3A, Genome-linked protein VP6; Picornain 3C
 DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
 DE (EC 2.7.7.48)).

OS Swine vesicular disease virus (strain H/3 '76).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxID=12076;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89279274; PubMed=2543767;
 RA Inoue T., Suzuki T., Sekiguchi K.;

RT "The complete nucleotide sequence of swine vesicular disease virus.";
 RL J. Gen. Virol. 70:919-924(1989).
 CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.

CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 CC O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Glu-1-Gly bond in the
 CC poliovirus polypeptide. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS.
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: D00435; BAA0037.1;
 CC PIR: A1331; GNNYSH.

DR HSSP: P21404; 1D4M.
 DR MEROPS: C03.022; -;
 DR InterPro: IPR000199; Cys-protease-3C.

DR InterPro: IPR000338; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.

DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; RVV.

DR Pfam: PF00073; rhv.3.
 DR Pfam: PF00548; Cys-protease-3C.1.
 DR Pfam: PF00680; RNA_dep_RNA_pol.1.
 DR Pfam: PF00910; RNA_helicase.1.

DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 DR Polypeptide: Coat protein; Core protein; Transferase;
 KM RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.

KW CHAIN 2
 FT CHAIN 70 330
 FT CHAIN 331 568
 FT CHAIN 569 851
 FT CHAIN 852 1001
 FT CHAIN 1002 1100
 FT CHAIN 1101 1429
 FT CHAIN 1430 1518
 FT CHAIN 1519 1540
 FT CHAIN 1541 1723
 FT CHAIN 1724 2185

FT CHAIN 2185 2185
 FT LIPID 2
 FT ACT_SITE 1687 1687
 FT ACT_SITE 1701 1701
 FT SEQUENCE 2185 AA; 243164 MW; 5B2285D094B2B3B6 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 30; DB 1; Length 2185;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 11
 POLG_SVDVH STANDARD; PRT; 2185 AA.

AC P1300; 084794; 084795; 084796; 084797; 084798; 084799; 084800;
 AC 084801; 084802; 084803; 084804;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Genome polypeptide (Contains: Coat proteins VP1 TO VP4; Core proteins
 DE P2A TO P2C; P3A, Genome-linked protein VP6; Picornain 3C
 DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
 DE (EC 2.7.7.48)).

OS Swine vesicular disease virus (strain UKG/27/72).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.

OX NCBI_TaxID=12077;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90364770; PubMed=2168111;
 RA Seechurn P., Knowles N.J., McCauley J.W.;

RT "The complete nucleotide sequence of a pathogenic swine vesicular
 RT disease virus.";
 RL Virus Res. 16:255-274(1990).

CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.

CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 CC O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Glu-1-Gly bond in the
 CC poliovirus polypeptide. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS.
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.

CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54521; CAA38377.1; -
 DR PIR: A30061; GNMISV.
 DR PIR: S11670; S11670.
 DR MEROPS: C03.011; -
 DR MEROPS: C03.022; -
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P2A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rnv.
 DR Pfam: PF00073; Rnv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 DR PolyProtein: Coat protein; Core protein; Transferrase;
 KM RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.
 FT CHAIN 2 69
 FT CHAIN 70 330
 FT CHAIN 331 568
 FT CHAIN 369 851
 FT CHAIN 852 1001
 FT CHAIN 1002 1100
 FT CHAIN 1101 1429
 FT CHAIN 1430 1518
 FT CHAIN 1519 1540
 FT CHAIN 1541 1723
 FT CHAIN 1724 2185
 FT LIPID 2
 FT ACT SITE 1687 1687
 FT ACT SITE 1701 1701
 FT ACT SITE 2185 2185
 FT SEQUENCE 2185 AA; 243363 MW; C9B103052934E1B8 CRC64;
 SQ

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
 DB 1425 EALFOG 1430

RESULT 12
 ID POLG_EC06C STANDARD; PRT; 2191 AA.
 AC Q66474;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein [contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
 DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 OS Echovirus 6 (strain Charles).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxID=103913;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Rightland V.F., Gratsch T.E., Blackburn R.V.;
 RT "The genome of echovirus 6";
 RL Submitted (OCT-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1526-2191 FROM N.A.
 RX MEDLINE=94240818; PubMed=8184543;
 RA Gratsch T.E., Rightland V.F.;
 RT "Constitution of a recombinant cDNA of echovirus 6 that established a
 RT persistent in vitro infection";
 RL Virology 201:341-348(1994).
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the
 CC picornavirus polyprotein. In other picornavirus reactions Glu may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS
 CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U16283; AA65044.1; -
 DR HSSP: P21404; 1D4M.
 DR MEROPS: C03.011; -
 DR InterPro: IPR004004; Calici_pol_hel.
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rnv.
 DR Pfam: PF00073; Rnv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR PRINTS: PR00918; CALICIVIRUSNS.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KM Polyprotein; Coat protein; Core protein; Transferrase; Myristate;
 KM RNA-directed RNA polymerase; Hydrolase; Thiol protease.
 FT CHAIN 2 69
 FT CHAIN 70 330
 FT CHAIN 331 568
 FT CHAIN 369 857
 FT CHAIN 858 1007
 FT CHAIN 1008 1106
 FT CHAIN 1107 1435
 FT CHAIN 1436 1524
 FT CHAIN 1525 1546
 FT CHAIN 1547 1729
 FT CHAIN 1730 2191
 RP RNA-DIRECTED RNA POLYMERASE.

FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
 FT ACT_SITE 1693 1693 PROTEASE 3C (POTENTIAL).
 FT ACT_SITE 1707 1707 PROTEASE 3C (POTENTIAL).
 SQ SEQUENCE 2191 AA; 245275 MW; 47E0448B05716DB CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2191;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1431 EALFOG 1436

RESULT 13
 ID POLG_CXA16 STANDARD; PRT; 2193 AA.

AC 065900;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein (Contains: Coat protein VP4 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Core protein P2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)).
 DE Picornavirus A16 (strain G-10).
 OS Cosackievirus ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Enterovirus.

OC Enterovirus.
 NX NCBI_TaxID=69159;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94303216; PubMed=8030260;
 RA Poiry T., Hyplise T., Horsnell C., Kinnunen L., Hovi T., Stanway G.;
 RT "Molecular analysis of cosackievirus A16 reveals a new genetic group of enteroviruses.";
 RL Virology 202;982-987(1994).

CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS
 CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Glu-1-Gly bond in the poliovirus polyprotein. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U05876; AA50478.1; .
 DR HSSP: P03300; IPOV.
 DR MEROPS: C03.022; .
 DR MEROPS: C03.0PA; .
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; RNV.
 DR Pfam: PF00073; thv_3.
 DR Pfam: PF00548; Cys-protease-3C; 1.

DR Pfam: PF00680; RNA_dep_RNA_pol_1.
 DR Pfam: PF00910; RNA_helicase_1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KM Polyprotein; Coat protein; Core protein; Transferase; Hydrolyase; Thiol protease; Myristate.
 KW RNA-directed RNA polymerase; RNA-directed RNA polymerase; RNA-directed RNA polymerase.
 FT CHAIN 2 69
 FT CHAIN 70 323
 FT CHAIN 324 565
 FT CHAIN 566 862
 FT CHAIN 863 1012
 FT CHAIN 1013 1111
 FT CHAIN 1112 1440
 FT CHAIN 1441 1526
 FT CHAIN 1527 1548
 FT CHAIN 1549 1731
 FT CHAIN 1732 2193
 FT LIPID 2 2
 FT ACT_SITE 1695 1695
 FT ACT_SITE 1709 1709
 SQ SEQUENCE 2193 AA; 243209 MW; 04B3BC572A76E38 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1436 EALFOG 1441

RESULT 14
 ID POLG_EC09H STANDARD; PRT; 2193 AA.

AC 066849;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein (Contains: Coat protein VP4 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)).
 DE Echovirus 9 (strain H11).
 OS Enterovirus.
 OC Enterovirus.
 NX NCBI_TaxID=103915;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96435007; PubMed=8837893;
 RA Zimmerman H., Eggers H.J., Krus W., Nielsen-Salz B.;
 RT "Complete nucleotide sequence and biological properties of an infectious clone of prototype echovirus 9.";
 RL Virus Res. 39:311-319(1995).

CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE AT CERTAIN O/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE PROTEASES.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Glu-1-Gly bond in the poliovirus polyprotein. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the poliovirus polyprotein. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

DR Pfam: PF00548; Cys-protease-3C; 1.

```

CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC: VP1/PA2 IS
CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X84981; CAA59341.1; -.
DR HSSP: 091734; 1EVL.
DR InterPro: IPR004004; Calic1_pol_hel.
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P2A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rhv.
DR Pfam: PF00073; rhv. 3.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR PRINTS: PR00918; CALICVIRUSNS.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
DR PolyProtein: Coat protein; Core protein; Transferrase; Myristate;
KM RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2 69
FT CHAIN 70 330
FT CHAIN 331 569
FT CHAIN 570 859
FT CHAIN 860 1009
FT CHAIN 1010 1108
FT CHAIN 1109 1437
FT CHAIN 1438 1526
FT CHAIN 1527 1548
FT CHAIN 1549 1731
FT CHAIN 1732 2193
FT LIPID 2 2
FT ACT SITE 1695 1695
FT ACT_SITE 1709 1709
SQ SEQUENCE 2193 AA; 245056 MW; E6ADD3091293C64A CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 1433 EALFOG 1438

RESULT 15
POLG_EC12T STANDARD; PRT; 2193 AA.
AC 066575; 066576;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
DE P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28)
DE (Protease 3c) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].

```

```

OS Echovirus 12 (strain Travis).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=103909;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Wild type;
RX MEDLINE=95364006; PubMed=7637032;
RA Kraus W., Zimmermann H., Zimmermann A., Eggers H.J., Nielsen-Salz B.;
RT Infectious cDNA clones of echovirus 12 and a variant resistant
RT against the uncoating inhibitor rhodamine differ in seven amino
RT acids.
RL J. Virol. 69:5853-5858(1995).
CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
CC AT CERTAIN O/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
CC PROTEASES.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Gln may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-I-Gly bond in the
CC picornavirus polyprotein. In other picornavirus reactions Gln may
CC be substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC: VP1/PA2 IS
CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79047; CAA55650.1; -.
DR EMBL: X77708; CAA54783.1; -.
DR HSSP: P21404; 1DAM.
DR MEROPS: C03.011; -.
DR MEROPS: C03.022; -.
DR InterPro: IPR004004; Calic1_pol_hel.
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rhv.
DR Pfam: PF00073; rhv. 3.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR PRINTS: PR00918; CALICVIRUSNS.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
DR PolyProtein: Coat protein; Core protein; Transferrase; Myristate;
KM RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2 69
FT CHAIN 70 330
FT CHAIN 331 568
FT CHAIN 569 859
FT CHAIN 860 1009
FT CHAIN 1010 1108
SQ CORE PROTEIN P2B.

```


FT CHAIN 1109 1437 CORE PROTEIN P2C.
FT CHAIN 1438 1526 CORE PROTEIN P3A.
FT CHAIN 1527 1548 GENOME-LINKED PROTEIN VPg.
FT CHAIN 1549 1731 PICORNAIN 3C.
FT CHAIN 1732 2193 RNA-DIRECTED RNA POLYMERASE.
FT LIPID 2 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1695 1695 PROTEASE 3C (POTENTIAL).
FT ACT_SITE 1709 1709 PROTEASE 3C (POTENTIAL).
FT VARIANT 223 223 H -> Y (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 228 228 G -> S (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 376 376 I -> M (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 643 643 Y -> C (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 669 669 V -> A (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 725 725 V -> A (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 2094 2094 C -> R (IN RHODAMINE-RESISTANT VARIANT).
SQ SEQUENCE 2193 AA; 244803 MW; 1EA34E298F53130C CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFQG 6
Db 1433 EALFQG 1438

Search completed: May 20, 2003, 11:35:01
Job time : 2.14163 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 10:12:24 ; Search time 0.413981 Seconds
(without alignments)
2986.328 Million cell updates/sec

Title: US-10-081-408-6
Perfect score: 30
Sequence: 1 EALFQG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	287	12	0910X8
2	30	100.0	310	2	09RNM8
3	30	100.0	522	10	093VM4
4	30	100.0	757	12	085082
5	30	100.0	989	10	09SCM5
6	30	100.0	1866	3	08TFN4
7	30	100.0	1930	12	0910W1
8	30	100.0	2063	12	08VOC4
9	30	100.0	2063	12	08VOC3
10	30	100.0	2063	12	08VOC2
11	30	100.0	2171	12	08VOC2
12	30	100.0	2174	12	09PYB7
13	30	100.0	2174	12	09PYB6
14	30	100.0	2183	12	08V639
15	30	100.0	2184	12	09Q188
16	30	100.0	2184	12	09QAH1

17	30	100.0	2184	12	09YX3	09YX3 human coxa
18	30	100.0	2185	12	066338	066338 human coxa
19	30	100.0	2185	12	09PYF2	09PYF2 human coxa
20	30	100.0	2185	12	09E7C4	09E7C4 human coxa
21	30	100.0	2185	12	09E7C3	09E7C3 human coxa
22	30	100.0	2185	12	09E7C2	09E7C2 human coxa
23	30	100.0	2185	12	09E7F10	09E7F10 human coxa
24	30	100.0	2185	12	088445	088445 swine vesic
25	30	100.0	2187	12	09QSM9	09QSM9 human coxa
26	30	100.0	2187	12	09YLG5	09YLG5 human coxa
27	30	100.0	2188	12	0999Y5	0999Y5 human coxa
28	30	100.0	2189	12	08V635	08V635 human echov
29	30	100.0	2193	12	09WPU0	09WPU0 human enter
30	30	100.0	2193	12	09WPU0	09WPU0 human enter
31	30	100.0	2193	12	09ORL8	09ORL8 human enter
32	30	100.0	2193	12	09QF52	09QF52 human enter
33	30	100.0	2193	12	09QF31	09QF31 human enter
34	30	100.0	2193	12	09DHE1	09DHE1 human enter
35	30	100.0	2193	12	09DY02	09DY02 human enter
36	30	100.0	2193	12	09DY01	09DY01 human enter
37	30	100.0	2193	12	0993S1	0993S1 enterovirus
38	30	100.0	2193	12	098Y36	098Y36 enterovirus
39	30	100.0	2194	12	0914E1	0914E1 human echov
40	30	100.0	2194	12	0914E0	0914E0 human echov
41	30	100.0	2194	12	08V640	08V640 human echov
42	30	100.0	2197	12	091M09	091M09 a-2 plaque
43	30	100.0	2206	12	09DIF1	09DIF1 human polio
44	30	100.0	2206	12	084792	084792 human polio
45	30	100.0	2207	12	098595	098595 human polio

ALIGNMENTS

RESULT 1
ID 0910X8 PRELIMINARY; PRT; 287 AA.
AC 0910X8;
DT 01-DEC-2001 (Trembl, 19, Created)
DT 01-DEC-2001 (Trembl, 19, Last sequence update)
DT 01-MAR-2002 (Trembl, 20, Last annotation update)
DE Polypeptide (Fragment).
OS Human echovirus 30.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OX NCBI_Taxid:41846;
RN [1]
RP SEQUENCE FROM N.A.
RA Mascillo M., La Rosa G., Mariani C., Capobianchi M.R.,
RA Zucchiatti S., Cattani P., Manzara S., Fadda G., Comparcola D.,
RA Tica F.;
RT "A phylogenetic analysis of the two echovirus 30 isolated in Rome
RT (Italy) in 1997 from an outbreak of aseptic meningitis associated with
RT swimming pools."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ295214; CAC38992.1;
DR InterPro: IPR000605; RNA_helicase.
DR Pfam: PF00910; RNA_helicase; 1.
FT NON_TER
FT CHAIN 1
FT NON_TER 1
SQ SEQUENCE 287 AA; 32023 MW; F284551D94A298C CRC64;
Query Match 100.0%; Score 30; DB 12; Length 287;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EALFQG 6
Db 272 EALFQG 277
RESULT 2

```

Q9RNM8
ID Q9RNM8 PRELIMINARY: PRT: 310 AA.
AC Q9RNM8:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE UDP-N-acetylglucosaminylglucosamine reductase.
GN MURB.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ZM4;
RA Um H.W., Kang H.S.;
RT "Sequence analysis of 41E10 fosmid clone of Zymomonas mobilis.";
Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF119611; AAD53934.1;
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
DR Pfam; PF02873; MurB_C; 1.
SQ SEQUENCE 310 AA: 33905 MW: 7DCA8C285C06B36 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 2; Length 310;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 190 EALFOG 195

RESULT 3
Q93VM4 PRELIMINARY: PRT: 522 AA.
AC Q93VM4:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative protease.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carinici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene T8H10.70/AT3957470 (GI:6706418).";
Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA Bowser L., Carinici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene T8H10.70/AT3957470 (GI:6706418).";
Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY035059; AAK59564.1;
DR InterPro; IPR001431; Peptidase_M16.

```

```

DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; UNKNOWN.1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN.1.
SQ SEQUENCE 522 AA: 59556 MW: 81836BFCCE011A2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 10; Length 522;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 235 EALFOG 240

RESULT 4
Q85082 PRELIMINARY: PRT: 757 AA.
AC Q85082:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Polypeptide precursor (Fragment).
OS Human poliovirus 1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=12080;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=81273079; PubMed=6267593;
RA Semler B.L., Anderson C.W., Kitamura N., Rothberg P.G., Wismart W.L.,
RA Wimmer E.;
RT "Poliovirus replication proteins: RNA sequence encoding P3-1b and the
RT sites of proteolytic processing.";
Proc. Natl. Acad. Sci. U.S.A. 78:3464-3468(1981).
DR EMBL; M17494; AAA46911.1;
DR InterPro; IPR000199; Cys-protease-3C.
DR InterPro; IPR001205; RNA_pol_P3D.
DR Pfam; PF00548; Cys-protease-3C; 1.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR PRODOM; PD001125; Cys-protease-3C; 1.
FT NON_TER 1 1
FT CHAIN 6 92 POTENTIAL.
FT CHAIN 93 114 POTENTIAL.
FT CHAIN 115 296 POTENTIAL.
FT CHAIN 297 757 POTENTIAL.
SQ SEQUENCE 757 AA: 84827 MW: C1B86D48A9D49060 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 12; Length 757;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1 EALFOG 6

RESULT 5
Q9SCM5 PRELIMINARY: PRT: 989 AA.
AC Q9SCM5:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Protease-like protein.
GN T8H10.70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]

```

RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Borkova D., Ansgorge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Quelet F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL33248; CAB6104.1; -;
 DR InterPro: IPR001431; Peptidase_M16.
 DR InterPro: IPR000169; SHProl_acsite.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; UNKNOWN; 1.
 SO SEQUENCE 989 AA; 113070 MW; E12CF684FE14E96 CRC64;

Query Match 100.0%; Score 30; DB 10; Length 989;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 235 EALFOG 240

RESULT 6
 O8TFN4 PRELIMINARY; PRT; 1866 AA.
 AC O8TFN4;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Chitin synthase C.
 GN CHSC.
 OS Colletotrichum graminicola (Anthracnose fungus) (Glomerella
 OS graminicola).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetes Incertae sedis; Phyllachorales; Phyllachoraceae;
 OC Glomerella.
 OX NCBI_Taxid=31870;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Werner S., Deising H.B.;
 RT "A csm-like chitin synthase of the maize pathogen Colletotrichum
 RT graminicola."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY053447; AAL23719.1; -;
 SO SEQUENCE 1866 AA; 207182 MW; 9126E926330DE1BF CRC64;

Query Match 100.0%; Score 30; DB 3; Length 1866;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 1145 EALFOG 1150

RESULT 7
 O91OW1 PRELIMINARY; PRT; 1930 AA.
 AC O91OW1;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Polyprotein (Fragment).
 OS Human echovirus 30.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_Taxid=41846;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-/DR/ROMA97;

RA Muscillo M., La Rosa G., Marianielli C., Capobianchi M.R.,
 RA Zaniratti S., Cattani P., Manzara S., Fadda G., Comparcola D.,
 RA Ticca F.;
 RT "A phylogenetic analysis of the two echovirus 30 isolated in Rome
 RT (Italy) in 1997 from an outbreak of aseptic meningitis associated with
 RT swimming pools."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ295172; CAC39011.1; -;
 DR MEROPS: C03.011; -;
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; Rhv.
 DR InterPro: IPR001205; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_p3b.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF00073; rhv; 3.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 FT CHAIN 1
 FT CHAIN 70
 FT CHAIN 331
 FT CHAIN 331
 FT CHAIN 569
 FT CHAIN 864
 FT CHAIN 1000
 FT CHAIN 1099
 FT CHAIN 1439
 FT CHAIN 1528
 FT CHAIN 1550
 FT CHAIN 1733
 FT CHAIN 1930
 FT NON_TER 1930
 SO SEQUENCE 1930 AA; 214968 MW; 559CFD7CC4F52FB2 CRC64;

Query Match 100.0%; Score 30; DB 12; Length 1930;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 1434 EALFOG 1439

RESULT 8
 O8VOC4 PRELIMINARY; PRT; 2063 AA.
 AC O8VOC4;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Polyprotein (Fragment).
 OS Human poliovirus 1.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_Taxid=12080;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-3788ALB96;
 MEDLINE-21635047; PubMed-11773144;
 RA Marturano J., Fiore L.,
 RT "Investigation of the Presence of Recombinant Polioviruses in the Hit
 RT Population in Albania during the 1996 Outbreak."
 RL J. Clin. Microbiol. 40:316-317(2002).
 DR EMBL: AY056701; AAL33795.1; -;
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR000345; CysC_heme_bind.

```
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR001676; Rnv.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF00073; Rnv; 3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
FT CHAIN 1 >69
FT CHAIN 70 >339 VP4.
FT CHAIN 340 >825 VP2.
FT CHAIN 826 >1127 VP3.
FT CHAIN 1128 >1276 VP1.
FT CHAIN 1277 >1403 PROTEASE P2-A.
FT CHAIN 1404 >1659 P2-B.
FT CHAIN 1670 >1789 P2-C.
FT CHAIN 1790 >1811 P3-A.
FT CHAIN 1812 >2001 VPG.
FT CHAIN 2002 >2063 PROTEASE P3-C.
FT CHAIN 2063 >2063 POLYMERASE P3-D.
SQ SEQUENCE 2063 AA; 229822 MW; 9ED5EA03BED37A4D CRC64;
```

Query Match 100.0%; Score 30; DB 12; Length 2063;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFQG 6
111111
DB 1447 EALFQG 1452

```
RESULT 9
Q8VOC3 PRELIMINARY; PRT; 2063 AA.
AC Q8VOC3;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Polypeptide (Fragment).
OS Human poliovirus 1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=12080;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3914ALB96;
RX MEDLINE=21635047; PubMed=11773144;
RA Marturano J., Fiore L.;
RT "Investigation in Albania during the 1996 outbreak.";
RL J. Clin. Microbiol. 40:316-317(2002).
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR000345; Cys-protease-3C.
DR InterPro: IPR003138; Cys-protease-3C.
DR InterPro: IPR000081; Pico_P1A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR001676; Rnv.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
```

```
DR Pfam: PF00073; Rnv; 3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
FT CHAIN 1 >69
FT CHAIN 70 >339 VP4.
FT CHAIN 340 >825 VP2.
FT CHAIN 826 >1127 VP3.
FT CHAIN 1128 >1276 VP1.
FT CHAIN 1277 >1403 PROTEASE P2-A.
FT CHAIN 1404 >1659 P2-B.
FT CHAIN 1670 >1789 P2-C.
FT CHAIN 1790 >1811 P3-A.
FT CHAIN 1812 >2001 VPG.
FT CHAIN 2002 >2063 PROTEASE P3-C.
FT CHAIN 2063 >2063 POLYMERASE P3-D.
SQ SEQUENCE 2063 AA; 229836 MW; A2ED9A21EDF22 CRC64;
```

Query Match 100.0%; Score 30; DB 12; Length 2063;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFQG 6
111111
DB 1447 EALFQG 1452

```
RESULT 10
Q8VOC2 PRELIMINARY; PRT; 2063 AA.
AC Q8VOC2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Polypeptide (Fragment).
OS Human poliovirus 1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=12080;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4019ALB96;
RX MEDLINE=21635047; PubMed=11773144;
RA Marturano J., Fiore L.;
RT "Investigation in Albania during the 1996 outbreak.";
RL J. Clin. Microbiol. 40:316-317(2002).
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR000345; Cys-protease-3C.
DR InterPro: IPR003138; Cys-protease-3C.
DR InterPro: IPR000081; Pico_P1A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR001676; Rnv.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF00073; Rnv; 3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
FT CHAIN 1 >69
FT CHAIN 70 >339 VP4.
FT CHAIN 340 >825 VP2.
FT CHAIN 826 >1127 VP3.
```

```

FT CHAIN 826 >1137 VP1.
FT CHAIN 1128 >1276 PROTEINASE P2-A.
FT CHAIN 1277 >1403 P2-B.
FT CHAIN 1404 >1669 P2-C.
FT CHAIN 1670 >1789 P3-A.
FT CHAIN 1790 >1811 VP.
FT CHAIN 1812 >2001 PROTEINASE P3-C.
FT CHAIN 2002 >2063 POLYMERASE P3-D.
FT NON-TER 2063 2063
SQ SEQUENCE 2063 AA: 229820 MW: 958FB9141805F2A1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 12; Length 2063;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1447 EALFOG 1452

RESULT 11
Q8QU26 PRELIMINARY; PRT; 2171 AA.
AC Q8QU26;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Polyprotein.
OS Porcine enterovirus 10.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=106967;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JP 54;
RA Krumholz A., Dauber M., Henke A., Birch-Hirschfeld E., Knowles N.J.,
RA Stelzner A., Zell R.;
RT "Sequencing of porcine enterovirus (PEV) groups II and III reveals
RT unique features of both virus groups."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF363455; AAM00276.1; -
SQ SEQUENCE 2171 AA: 240751 MW: 789FE0C3A7FD378 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 12; Length 2171;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 1412 EALFOG 1417

RESULT 12
Q9PYB7 PRELIMINARY; PRT; 2174 AA.
AC Q9PYB7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Polypeptide.
OS Bovine enterovirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=12064;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K2577;
RA MEDLINE=99429271; PubMed=10501163;
RA McCarthy F.M., Smith G.A., Mattick J.S.;
RT "Molecular characterisation of Australian bovine enteroviruses."
RL Vet. Microbiol. 68:71-81(1999).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-K2577;
RA McCarthy F.M., Smith G.A., Mattick J.S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF123432; AAF05693.1; -
DR HSSP; P12915; 1BEV.
DR InterPro; IPR004004; Calic1_pol_hel.
DR InterPro; IPR000199; Cys-protease-3C.
DR InterPro; IPR003138; Pico_P1A.
DR InterPro; IPR000081; Pico_P2A.
DR InterPro; IPR002527; Pico_P2B.
DR InterPro; IPR001676; Rhv_P2B.
DR InterPro; IPR006055; RNA_helicase.
DR InterPro; IPR001205; RNA_pol_P3D.
DR Pfam; PF00548; Cys-protease-3C; 1.
DR Pfam; PF02226; Pico_P1A; 1.
DR Pfam; PF00947; Pico_P2A; 1.
DR Pfam; PF01552; Pico_P2B; 1.
DR Pfam; PF00073; Rhv_P3.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSNS.
DR ProDom; PD001125; Cys-protease-3C; 1.
DR ProDom; PD001274; Pico_P2B; 1.
DR ProDom; PD001306; Pico_P2A; 1.
FT CHAIN 1
FT CHAIN 70 71 VP4.
FT CHAIN 318 559 VP2.
FT CHAIN 560 839 VP3.
FT CHAIN 840 989 VP1.
FT CHAIN 990 1088 PROTEINASE 2A.
FT CHAIN 1089 1418 2B.
FT CHAIN 1419 1507 2C.
FT CHAIN 1508 1530 3A.
FT CHAIN 1531 1713 VP.
FT CHAIN 1714 2174 PROTEINASE 3C.
SQ SEQUENCE 2174 AA: 242178 MW: 3D4787508B05288 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 12; Length 2174;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1414 EALFOG 1419

RESULT 13
Q9PYB6 PRELIMINARY; PRT; 2174 AA.
AC Q9PYB6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Polypeptide.
OS Bovine enterovirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=12064;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SL305;
RA MEDLINE=99429271; PubMed=10501163;
RA McCarthy F.M., Smith G.A., Mattick J.S.;
RT "Molecular characterisation of Australian bovine enteroviruses."
RL Vet. Microbiol. 68:71-81(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SL305;
RA McCarthy F.M., Smith G.A., Mattick J.S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF123433; AAF05694.1; -
DR HSSP; P12915; 1BEV.
DR InterPro; IPR004004; Calic1_pol_hel.

```

DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P2A.
 DR InterPro: IPR000081; Pico_P2B.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; Rhv.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF00073; Rhv; 3.
 DR Pfam: PF00580; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR PRINTS: PR00918; CALICIVIRUS.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 FT CHAIN 1
 FT CHAIN 1
 FT CHAIN 70
 FT CHAIN 317
 FT CHAIN 318
 FT CHAIN 559
 FT CHAIN 560
 FT CHAIN 839
 FT CHAIN 840
 FT CHAIN 989
 FT CHAIN 990
 FT CHAIN 1088
 FT CHAIN 1089
 FT CHAIN 1418
 FT CHAIN 1419
 FT CHAIN 1507
 FT CHAIN 1508
 FT CHAIN 1530
 FT CHAIN 1531
 FT CHAIN 1713
 FT CHAIN 1714
 FT CHAIN 2174
 SO SEQUENCE 2174 AA; 242286 MW; 9AADC6487649A2BC CRC64;

Query Match
 Best Local Similarity 100.0%; Score 30; DB 12; Length 2174;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 1414 EALFOG 1419

RESULT 14
 O8V639 PRELIMINARY; PRT; 2183 AA.
 AC O8V639;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Polypeptide.
 OS Human coxsackievirus B4.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OC NCBI_TaxID=12073;
 RN NCBI_TaxID=12073;
 RP SEQUENCE FROM N.A.
 RC STRAIN=E2 VARIANT;
 RA Lindberg A.M., Andersson P.;
 RT "Molecular analysis of a coxsackievirus B4 strain E2 variant."
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF131939; AAL37156.1; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; Rhv.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF00073; Rhv; 3.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.

DR Pfam: PF00910; RNA_helicase; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 DR SMART: SM00382; AAA; 1.
 SO SEQUENCE 2183 AA; 244045 MW; BC34D472E37953D CRC64;

Query Match
 Best Local Similarity 100.0%; Score 30; DB 12; Length 2183;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 1423 EALFOG 1428

RESULT 15
 O9OL88 PRELIMINARY; PRT; 2184 AA.
 AC O9OL88;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Polypeptide.
 OS Human coxsackievirus B6.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OC NCBI_TaxID=74561;
 RN NCBI_TaxID=74561;
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCHMITT;
 RX MEDLINE-99432326; PubMed-10500285;
 RA Martino T.A., Tellier R., Petric M., Irwin D.M., Afshar A., Liu P.P.;
 RT "The complete consensus sequence of coxsackievirus B6 and generation
 of infectious clones by long Rr-PCR."
 RL Virus Res. 64:77-86(1999).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCHMITT;
 RA Martino T.A., Tellier R., Petric M., Irwin D.M., Afshar A., Liu P.P.;
 RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF105342; AAF12719.1; -
 DR HSP: P21404; 1D4M.
 DR MEROPS: C03.022; -
 DR InterPro: IPR004004; Calic1_pol_hel.
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; Rhv.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF00073; Rhv; 3.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR PRINTS: PR00918; CALICIVIRUS.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 FT CHAIN 1
 FT CHAIN 69
 FT CHAIN 70
 FT CHAIN 330
 FT CHAIN 331
 FT CHAIN 568
 FT CHAIN 569
 FT CHAIN 850
 FT CHAIN 851
 FT CHAIN 1000
 FT CHAIN 1001
 FT CHAIN 1099
 FT CHAIN 1100
 FT CHAIN 1429
 FT CHAIN 1429
 FT CHAIN 1517
 FT CHAIN 1518
 FT CHAIN 1539
 FT CHAIN 1540
 FT CHAIN 1722
 FT CHAIN P3-C.

FT CHAIN 1723 2184 P3-D, 48B8268B1C523AD8 CRC64;
SQ SEQUENCE 2184 AA; 243253 MM; 100.0%; Score 30; DB 12; Length 2184;
Query Match
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EALFQG 6
11111
Db 1424 EALFQG 1429

Search completed: May 20, 2003, 11:37:49
Job time : 2.41398 secs